

국문학 내 문학사회학과 멀리서 읽기*

— 새로운 검열연구를 위한 길마중

李載然* · 鄭瑜旻**

- | | |
|-----------------------------|-----------------------|
| I. 영미문학의 멀리서 읽기와 문학사회학 | V. 정량분석, 매체연구, 검열연구 |
| II. 골드만, 김현과 소설사회학 | VI. 근현대 한국 문학의 멀리서 읽기 |
| III. 에스카르피, 이선영, 권영민과 문학사회학 | VII. 覆字의 멀리서 읽기 시론 |
| IV. 김영민과 매체 중심의 문학사회학 | VIII. 결론을 대신하여 |

● 국문초록

최근 미국의 영문학에서는 프랑코 모레티가 제창한 ‘멀리서 읽기’(distant reading)라는 문학내 정량분석을, 기존의 문학사회학적 전통과 디지털 인문학을 잇는 연결고리로 위치시키는 작업이 한창이다. 이러한 해외의 동향을 참조하여, 본 논문은 한국문학 내 문학사회학의 계보 속에 디지털 문학 방법론의 위상을 찾고 앞으로 전개할 새로운 연구 방향을 모색하고자 한다. 그 모색은 본 논문에서 크게 두 부분으로 구성되어 있다. 우선, 1980년대에 전개된 문학사회학에 관한 논의가 통계기반의 문학생산 분석, 신문 및 잡지 매체 연구 및 검열 연구를 거쳐, 디지털 문학연구로 수렴되는 과정을 파악한다. 이 수렴의 연장선상에서, 디지털 문학연구의 텍스트 마이닝 방법 (고빈

* 본 논문은, 2019년 필자들이 시카고 대학의 최경희 선생님과 함께 Modern Language Association의 연례학술대회에 “Text-Mining the Censor's Examination Copy: The Unpublished Collection of Poems by Shim Hun”이라는 연구를 발표하면서 얻은 문제의식을 바탕으로 쓴 글이다. 디지털 문학방법론을 검열연구에 적용해 보자는 화두를 던져주신 최경희 선생님께 감사드린다.

** UNIST(울산과학기술원) 부교수

*** 한남대학교 조교수

도어 분석, 공기어분석, 워드 임베딩 등)을 실제 검열연구에 적용하여 OO나 XX와 같은 복자(후세지, 삭제된 글자 대신 표시한 OO, XX와 같은 기호)의 전후 문맥에서 반복되는 의미의 맥락을 살펴볼 것이다. 뒷부분의 시도는 아직 실험적인 단계이고, 논문은 국문학 내 문학사회학적 전개과정에 더 초점이 맞춰져 있지만 이와 같은 시도를 통해 국문학에서 새롭게 개진할 연구의 향방을 가늠해 보고자 한다.

주제어 : 멀리서 읽기, 문학사회학, 매체론, 검열연구, 텍스트 마이닝, 한국 근대문학, 디지털 문학연구

I. 영미문학의 멀리서 읽기와 문학사회학

프랑코 모레티의 ‘멀리서 읽기’(distant reading)는 정량 분석에 대한 미국 내 영문학의 관심을 불러일으켰다.¹⁾ 이 새로운 읽기를 범박하게 요약하면, 1%의 문학 정전에 집중하면서 잊게 된 99%의 작품을 문학사 안으로 들여와 다시 연구하자는 제안이다. 1940년대 신비평의 등장 이후 꼼꼼히 읽기(close reading)는 작품 분석의 주된 방법이였다. 멀리서 읽기는 이 방식과 거리를 두면서, 그간 문학 방법론상 주변의 위치에 있었던 정량분석의 가치를 提高하는 제안이기도 하다.

이렇게 재조정된 초점은 멀리서 읽기의 이론과 적용에 관해 치열한 논쟁을 촉발했다. 이 방법을 부정하는 학자들은 작품 읽기는 정량 분석의 대상이 될 수 없다는 전제에서, 계량화가 작품 읽기의 즐거움을 해친다고 주장한다.²⁾ 문학은 데이터로 환원될 수 없지만,³⁾ 설사 전산 분석을 통해 문학사에 접근한다고 하더라도 그것은 프로그램의 영역에서 의미 있을 뿐 해석이 아니며,⁴⁾ 모레티와 그 추종자들은 이 분석 결과를 마치 불변의 역사적 사실처럼 다룬다는 측면에서 비/몰역사적이라고 비판했다.⁵⁾ 한편, 멀리서 읽기의 비판적 지지자들은, 꼼꼼히 읽기를 오래된 정전 해석, 멀리서 읽기를 새롭지만 위험한 방법으로 나누는 이분법에 관해, 이전의 사회과학의 “패러다임 전쟁”을 연상시키는 등 무익하며, 오히려 이 두 읽기 방식은 서로의 방법론적 단점을 보완하는 측면에서 유효하다고 주장한다.⁶⁾ 나아가 그들은 나아가

-
- 1) 프랑코 모레티 저, 이재연 역, 『그래프, 지도, 나무』, 문학동네, 2020(영문판 2005); Franco Moretti, *Distant Reading*, London: Verso, 2013.
 - 2) Dora Zhang, “Literature and Money: Studying What You Love Without Being Exploited”, *Chronicles of Higher Education*, 2015.
 - 3) Stephen Marche, “Literature Is not Data: Against Digital Humanities,” *Los Angeles Review of Books*, 2012.
 - 4) Timothy Brennan, “The Digital Humanities Bust”, *The Chronicle Review*, 2017, B12~B14.
 - 5) Katherine Bode, “The Equivalence of ‘Close’ and ‘Distant’ Reading; or, Toward a New Object for Data-Rich Literary History”, *MLQ*, 2017, 79~81면.
 - 6) J. Berenike Herrmann, “In a Test Bed with Kafka. Introducing a Mixed-Method Approach to Digital Stylistics”, *DHQ* Vol. 11, No. 4, 2017; S. Jänicke, G. Franzini, M. F. Cheema1 and G. Scheuermann, “On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges,” *Eurographics Conference on Visualization (EuroVis)*, 2015, <https://www.informatik>.

멀리서 읽기가 계량역사학, 사회지리학, 생물학적 진화론과 같은, 그간 문학적 방법론에서 다루지 않은 방법들을 문학연구 안으로 끌고 들어옴으로써 문학연구의 지평을 열었다는 찬사도 함께 덧붙였다.⁷⁾

멀리서 읽기가 촉발한 20년 가까운 논쟁 후, 이 새로운 읽기는 미국 내 영미문학에서 하나의 방법론으로 자리잡힌 듯하다. 특히 멀리서 읽기가 환기하는 디지털 인문학(digital humanities 혹은 디지털 문학연구 digital literary studies 혹은 문화분석학 cultural analytics)을 일반적 문학연구 내 분과학문으로 정립하려는 다양한 시도들이 나타났다. 예를 들어 영문학 내 디지털 문학연구를 선도하고 있는 테드 언더우드(Ted Underwood)는 ‘멀리서 읽기’라는 용어는 프랑코 모레티가 새로 만든 것이 맞지만, 그 방법론의 근간이 된 통계분석이나 거시분석은 이전의 문학연구 방법에 이미 있었던 것이라고 주장한다.⁸⁾ 특히 그는, 물질주의적 관점에서 문화변동의 거시적 흐름을 파악한 레이몬드 윌리엄스(Raymond Williams, 1921~1988)의 『키워드』(Keywords, 1976)와 같은 마르크스주의 문화주의 저작, 사회과학적 질문조사방법을 통해 얻는 계량화된 독자의 감상을 기반으로 로맨스 장르의 문학적 특성을 살펴본 제니스 레드웨이(Janice Radway)의 『로맨스 읽기』(Reading the Romance, 1991)와 같은 문학사회학 연구 등을 정량분석의 전범으로 삼고 모레티의 멀리서 읽기와 연결시켰다. 비슷한 맥락에서 부르마(Rachel Buurma)와 헤프먼(Laura Heffernan)은, 보통 로베르토 부사(Roberto Busa, 1913~2011)로 인식되는 인문학 내 전산분석의 기원⁹⁾을 1930년대 버클리 대학에서 시를 가르쳤던 조세핀 마일즈(Josephine Miles, 1911~1985)로 재위치시켰다.¹⁰⁾ 마일즈는 컴퓨터를 활용하여 던(Donne), 밀턴(Milton), 포프(Pope)

uni-leipzig.de/~stjaenicke/Survey.pdf.

- 7) 알베르토 피아자, 『그래프, 지도, 나무』의 발문, 「가까운 영역에서의 진화」 참조; 한국학계에서 모레티를 수용한 과정은, 김용수, 「세계문화과 디지털 인문학 방법론-한국 학계의 모레티 연구」, 『비평과 이론』 24.3, 2019, 59~78면 참조.
- 8) Ted Underwood, “A Genealogy of Distant Reading”, *Digital Humanities Quarterly* Vol. 11, No. 2, 2017.
- 9) 부사는 1949년부터 1974년 동안 IBM 프로그램을 활용하여 토마스 아퀴나스의 중세 라틴어 저작 색인을 만들었다.
- 10) Rachel Sagner Buurma and Laura Heffernan, “Search and Replace: Josephine Miles and the Origins of Distant Reading”, *Modernism/modernity Print Plus* Vol. 3, Cycle 1, 2018, <https://modernismmodernity.org/forums/posts/search-and-replace>.

와 같은 낭만주의 시인들이 쓴 형용사를 색인으로 만들고 빈도를 구하여 이를 해석한 여성 문학자다. 저자들이 마일즈에 주목한 이유는, “문학사에 대한 정량적 혹은 경험적 접근”이라는 측면에서 부사보다 더 가깝고 또한 이것이 전산비평이라 불리는, 멀리서 읽기가 수렴한 한 방법론의 기원이 된다고 판단했기 때문이다. 이렇듯 새로운 문학연구사 기술은 기존의 형식주의, 문학사회학, 책의 역사, 디지털 인문학 등 이전에는 관련성이 적었던 여러 분과학문을 문학사 내 정량분석이라는 큰 흐름으로 잇고 있다.¹¹⁾

이 글은 우선, 앞서 간단하게 살펴본 미국 내 영문학의 연구사적 흐름을 참조하여 1980년대 이후의 한국문학에서 진행된 문학사회학의 계보를 훑어볼 것이다. 특히 정량분석을 사용한 문학생산 분석, 신문 및 잡지 매체 연구, 또한 검열 연구 등의 문학사회학적 흐름 속에 디지털 문학연구라는 새로운 방법론을 위치시키고자 한다. 이런 시도를 통하여 현재의 학문 구도를 형성한 과거의 학문적 문맥을 이해하는 한편, 디지털 연구에서 강조된 전산분석을 1930년대에 검열된 텍스트에 적용하여 앞으로 문학사회학에서 새롭게 등장할 연구를 예비해보고자 한다.

II. 골드만, 김현과 소설사회학

미국의 문학사회학자인 제임스 잉글리쉬(James F. English)는 최근의 “문학사회학 이후의 문학사회학”이라는 논문에서 영미의 문학사회학 연구의 계보를 간략하게 짚었다.¹²⁾ 그는 새로운 사회학 연구의 미래를 문학 연구 안에서 찾겠다는 희망을 품고

11) 예를 들면, 책의 역사라는 분과학문에서, 중요한 업적으로 인정받은 프리야 조시(Priya Joshi)의 저작, 『다른 나라에서』(*In another country*)는 최근에는 전산분석 혹은 디지털 인문학 연구 방법론의 전사(pre-history)로 언급되고 있다. Mike Frangos, “The End of Literature: Machine Reading and Amitav Ghosh’s *The Calcutta Chromosome*”, *Digital Humanities Quarterly*, Vol. 7, No. 1, 2013; 조시의 저서는 식민지 인도로 수입된 영국 서적의 통계 분석을 통해, 인도 독자들이 왜 영어로 된 로맨스 소설을 즐겨 읽었나를 분석한 책이다. Priya Joshi, *In Another Country: Colonialism, Culture, and the English Novel in India*, New York: Columbia University Press, 2002.

12) James F. English, “Everywhere and Nowhere: Sociology of Literature after the ‘Sociology of

1970년대 영국의 신좌파부터 2000년대 프랑코 모레티의 연구까지를 연결하여 살펴 보았다. 리처드 호가드, 스튜어트 홀 등, 맑시즘의 경제사회학적 관점을 문화 연구로 전환한 1970년대 버밍햄 학파는 영국의 레이몬드 윌리엄스, 프랑스의 루시앙 골드만(Lucien Goldmann, 1913~1970) 및 로베르 에스카르피(Robert Escarpit, 1918~2000) 등에 영향을 주었고, 이들에 의해 문학사회학이라는 분과학문이 확립되었다.

잉글리쉬에 따르면 문학사회학은 아래와 같은 몇 가지의 학제간 미션을 추구한다. “텍스트의 개별 형식 속 사회적 의미를” 파악하고, “텍스트 생산[에 영향을 준] 사회적 집단을 참조하여 …… 그들의 문학적 실천”을 설명하며, “텍스트 유통과 소비가 불러온 사회적 효과”를 분석한다.¹³⁾ 이와 같은 텍스트를 대상으로 한 학제간 연구를 통해 문학사회학은 책의 역사라는 영역으로 확장된다. D.F 메켄지(D.F. McKenzie), 로저 샤프티에(Roger Chartier), 로버트 단턴(Robert Darton) 등은 텍스트 생산과 유통에 관여한 여러 사회집단을 분석하여 텍스트 해석에 반영한 학자들이다. 특히 메킨지는 “텍스트의 사회학”(sociology of texts)이라는 개념으로, 텍스트 생산에 있어서 반영되는 물질성을 해석한다. 모든 판본을 독립적인 텍스트로 이해하고 이 생산에 관여한 모든 물질적 조건들(종이질, 잉크, 폰트, 활자, 등등)을 고려하여 각각의 판본을 독립적으로 이해할 것을 주문한다.¹⁴⁾ 1980년대 문학사회학은, 이러한 책의 역사와의 연계와 더불어, 피에르 브루디외(Pierre Bourdieu 1930~2002)의 장(field) 이론(각기 다른 문화 자본에 기반한 예술 생산 집단이 구별짓기의 룰을 구조화하고 경쟁하는 장소로서의 예술장), 해석 주체로서 독자를 강조한 콘스탄츠 학파의 독자-반응 이론과 연계하며 확장하였다. 1990년대에는 탈식민주의, 퀴어 연구, 신역사주의 등등의 방계 학문 분야로 그 외연이 확장되었고, 2000년대 이후에는 프랑코 모레티와 파스칼 카사노바(Pascale Casanova)의 세계문학 논의, 즉 “지구화의 사회학”(sociology of globalization)과 같은 새로운 영역에서 문학사회학의 가능성이 타진되고 있다는 것이 잉글리쉬의 설명이다.

한편, 한국에서 서양의 문학사회학 계보를 체계화하고 이를 국문학에 적용하여

Literature”, *New Literary History* 41, 2010, pp. v~xxiii.

13) 잉글리쉬, 앞의 논문, viii면.

14) D. F. McKenzie, “The Book as an Expressive Form”, *Bibliography and the Sociology of Texts*, London: Cambridge, 1999, pp. 9~30.

살펴본 이 중에 불문학자 김현(1942~1990)이 있다. 『문학사회학』(1983)에서 보인 그의 문학연구사 기술 방식은 크게 두 가지로 요약할 수 있다. 우선 문학사회학적 기원을 잉글리쉬보다 훨씬 앞선 시기로 잡고 있고, 제창자의 전기적 기술, 이론의 초점, 영향관계 등을 상세하고 구체적으로 정리하고 있다. 둘째(이 입장이 더 중요한 데), 1970~1980년대의 여러 문학사회학 이론 중 루시앙 골드만의 “소설사회학”을 크게 중시하고 이를 에스카르피의 “문학사회학”보다 우위에 두었다.¹⁵⁾ 그의 저서는 외국 이론을 정리하여 한국 작품에 적용하고 있기는 하지만, 비평적 관점은 이론들의 정리보다 두각을 보이지는 못한다.

김현이 본 문학사회학의 시작은 18~19세기의 스탈 부인(Madame de Staël, 1766~1817)과 텐느(Hippolyte Adolphe Taine, 1828~1893)이다. 스탈 부인은, 문학의 미적 다양성을 한 나라의 풍토적 다양성으로 설명하였고 텐느는 이를 종족, 환경(유전적 및 자연적), 문화구축의 시기 등으로 확장하였다. 이러한 초기 이론을 더 정치하게 발전시킨 이들이 플레하노프나 루카치, 골드만 같은 맑시스트 비평가들이었다. 이들은 문학의 사회반영, 특히 계급투쟁 현장으로서 사회를 반영함을 주장했는데 특히 루카치는 이를 좀 더 섬세하게 다듬어 “장르의 사회학”으로 발전시켰다.¹⁶⁾ 사회의 근대화 과정에서 잃어버린 전근대의 서사시적 총체성을, 문제적 개인과 세계와의 대립을 통해서 찾아가는 장르가 소설이라는 그 유명한 총체성론은 여기에서 나왔다. 이를 골드만은 “소설사회학”이라는 일종의 구조주의 이론으로 변모시킨다. 핵심은 의미로서의 소설과 상품으로서의 소설 사이에 상동성이 존재한다는 것. 즉, 주인공이 세계와의 대립 속에 자신의 가치를 추구하는 과정은, 상품으로서의 소설이 자본주의 시장에서 자신의 가치를 추구하는 과정과 유사하다는 것이다. 골드만은, 타락한 세계에서 타락한 방식으로 주인공이 삶의 진정한 가치를 건져내는 방식은, 교환가치가 사용가치보다 우위에 서는 자본주의의 물화된 시장 속에서, 상품으로 소비되는 형식으로만 소설의 가치가 획득되는 것과 마찬가지로 주장하였다.¹⁷⁾ 김현은 이와 같은 소설사회학 이론가들은 심도 있게 설명하지만 에스카르피와 같은

15) 김현, 『한국문학의 위상(1983)/문학사회학』, 문학과지성사, 1991, 302면.

16) 김현, 앞의 책, 251면.

17) 골드만 저, 오생근 역, 「소설사회학을 위한 서론」, 유종호 편, 『문학 예술과 사회 상황』, 민음사, 1979, 206면; 김현 앞의 책, 268면.

계량주의적 문학사회학자는 간략히 언급하고, 아예 브루디외의 예술장 이론은 건너 뛰고 있다. 즉 그는 예술 취향의 집단성이나 그러한 취향이 구조화된 문학생산 관계, 그리고 더 큰 의미에서 문학의 제도성이나 물질성보다는, 소설이라는 장르가 내재한 시대에 대한 저항정신에 좀 더 천착한 것으로 보인다.¹⁸⁾

Ⅲ. 에스카르피, 이선영, 권영민과 문학사회학

김현의 『문학사회학』 저서에는 비중이 작게 다루졌지만, 에스카르피의 저작은 1980년대 초중반 국내에서 활발하게 번역되었다. 1983년 김현의 『문학사회학』이 출간된 그해 에스카르피의 『문학의 사회학』이 출간되었고, 2년 뒤 『책의 혁명』이 출간되었다.¹⁹⁾ 『문학의 사회학』 붙어 원본(*Sociologie de la Littérature*)이 1958년에 나왔으니 불문학자 김현도 익히 이 책에 관해 알고 있었을 것이다. 프랑스 보르도 대학의 비교문학 교수였던 저자는, 1964년 리처드 호가트(Richard Hoggart, 1918~2014)가 버밍햄 대학에 현대문화연구소를 세우기도 전인 1960년 보르도에 문학사실 사회학 연구소(Center for Sociology of Literary Facts)를 세우고 통계를 활용한 문학사회학 연구를 시작하였다. 이 문학사실을 구성하는 세 가지 요소는, 서적, 독서, 문학으로, 구체적으로 말하자면 물질적 대상으로서의 서적, 예술적 및 기능적 목적을 포함하는 독서, 그리고 창작자와 소비자로서의 독자가 맺는 관계로서의 문학이다.²⁰⁾

에스카르피는 이 중, 작가-독자-독서를 매개하는 물질적 방식이자 형식이었던 서적을 가장 중요시했고, 앞서 질베르 뮈리(Gilbert Mury)가 고민했던 “서적 사회학은 가능한가?”라는 질문에 답을 구할 수 있는 자신만의 방식을 찾았다. 그는 작가, 독자, 서적중개인, 사서에게 양케이트를 실시하여 문학사실을 구하는 방식은 객관적이지

18) 이러한 김현의 비평적 방향성은, 당시의 군부독재나 광주민주화운동과 같은 역사 사회적 사건들에 의해 크게 영향받은 것으로 보인다. 그러나 필자의 연구가 부족하여 아직 그 맥락을 풀어내는 데까지 나아가지는 못했다.

19) 로베르 에스카르피 저, 민희식, 민병식 공역, 『문학의 사회학』, 을유문화사, 1983; 임문영 역, 『책의 혁명』, 보성사, 1985.

20) 에스카르피, 앞의 책, 29~37면.

않다고 보고 서적을 정량적으로 분석하여 객관적 데이터를 얻고자 했다. 유네스코 유럽위원회는 1950년대 초반부터 1960년대(1966년에는 에스카르피 자신도 참여) 사이 유럽의 서적출판, 수출입, 번역 등에 관한 통계를 작성하였는데 그는 이를 활용하여 프랑스 사례를 중심으로 비교적 시각을 확보하였다. 더불어 작가리스트는, 출처가 다른 문학 관련 자료—『작품사전』, 복각서나 번역서의 카탈로그, 잡지의 목차 등등—에서 작가 정보를 얻어 백과사전적 성격을 띤 목록(예, 『국민인명사전 *Dictionary of National Biography*』)과 비교, 대조하는 방식으로 구했다. 이 결과 유식자 그룹과 그렇지 않은 작가를 포함하여, “1490년부터 1900년 사이에 태어난 937명의 프랑스 작가” 표본이 도출되었다.²¹⁾

에스카르피는 이 표본에서 작가의 연령, 등단 연도, 장르별 작가 수 등을 구했다. 특히 장르별 작가 수를 전체 작가 수로 나누어 시계열로 늘어놓았는데 이를 통해 문학사의 중요한 발견을 할 수 있었다.

예를 들면, 하나의 문학적 장르—엘리자베트 시대의 비극, 고전 비극, 제18세기 영국의 사실 소설, 로만주의의 운동들—의 ‘생명’은 대략 30내지 35년, 즉 인간의 생명의 반쯤에 해당한다. …… 우리는 …… 문학 인구 [작가] 전체에 대한 소설가, 시인, 극작가, 각종 산문작가의 상호적 비율[다방면에서의 작가는 각종의 장르에 한 사람을 몇 차례든 계산에 넣기로 하고]을 나타내는 곡선을 서로 겹쳐 놓았다. 상당히 분명하게 알게 된 것은, 도표는 70년마다 근본적으로 바뀐다는 것, 그리고 이리이러한 장르가 다른 장르를 압도하거나 혹은 한때 모습을 감추거나 함에 따라 부분적으로는 35년마다 바뀐다는 것이다.²²⁾

이 발견은, 모레티가 『그래프, 지도, 나무』(2005)에서 주장했던 한 장르의 세대론, 혹은 30년 주기론보다 거의 반세기 앞서서 제기된 주장이다.²³⁾ 물론 에스카르피는 한 장르를 생산하는 작가의 성쇠를, 모레티는 한 장르를 소비하는 독자의 성쇠를 본 차이점이 있지만, 장르의 주기론은 에스카르피에 의해 거의 50년 전에 제기되었다는 점에서 의의가 있다.

21) 에스카르피, 앞의 책, 43면.

22) 에스카르피, 앞의 책, 58면.

23) 프랑코 모레티, 『그래프, 지도, 나무』, 28~34면.

에스카르피는 문학의 생산 측면과 아울러 분배에도 관심이 있었다. 그는 한 나라당 작가수, 서적 출판량, 번역량, 인구수 등을 조사하였고 또한 이를 세계의 주요 언어권 사용자/사용권역 속에서 파악하여 문학의 거시적 분배 회로를 파악하였다. 그는 이를 “문학상의 고기압 지대”와 “저기압 지대”라는 신조어로 표현하였는데, 전자는 영어, 중국어, 러시아어 등 세계 주요 언어의 사용자 지대이고 문화 수준이 높아 많은 작가가 독자들의 독서욕을 쉽게 만족하게 하는 블록이다.²⁴⁾ 따라서 이런 지대에서는 서적 출판이 빠르게 포화 상태가 된다. 이에 출판업자들은 번역으로 출판의 활로를 찾는다. 번역을 통해 문학상의 저기압 지대로 서적이 흘러 들어가, 그 지역에서는 쉽게 채워지지 않는 독서욕을 채우게 된다고 에스카르피는 보았다. 지구적인 시각에서 본 문학의 중심부와 주변부의 차이, 번역을 통한 두 지역의 소통과 흐름 등, 에스카르피의 지구적 분배 회로는 최근 세계문학론에서 자주 언급된 파스칼 카사노바의 “세계문예공화국”을 연상시킨다.²⁵⁾ 카사노바는 브루디외의 문학장 이론과 브로텔의 세계시스템을 결합하여 이러한 가상 문예공화국의 틀을 만들었지만, 에스카르피는 작가, 서적, 각국 인구에 관한 통계만으로 비슷한 모델을 40년 전에 이미 만들 수 있었던 것이다.

위와 같은 에스카르피식 문학사회학을 전개하려면 최소한 작가와 서적이거나 작품 출판의 연도별 목록이 필요하다. 그러나 한국문학의 경우 문학 생산 관련 목록과 이에 근거한 기초 통계 작성은 에스카르피의 영향을 받은 이선영에 의해 1990년대에야 작성되었다. 그렇지만 문학 연표 자체가 없었던 것은 아니었다. 예를 들어, 1975년 출간된 김병철의 『한국 근대번역문학사 연구』는 세세한 번역 서지 목록을 제공한다.²⁶⁾ 본격적인 정량분석이라고는 할 수 없지만, 앞서 잉글리쉬가 주장한 문학사회학의 한 부분, “텍스트 유통과 소비가 불러온 사회적 효과”의 입장에서 살펴볼 수 있는 마중물이 된다. 번역자이기도 했던 저자는 “이름조차 전해지지 않는 사람들이 …… 번역작품을 남기고 있다는 사실”을 보며 자신도 그들과 똑같은 운명에 처할지도 모른다는 위기감에 이 저서를 남기게 되었다고 적고 있다.²⁷⁾ 『번역문학사 연구』는 개화기 성서와 찬송가 번역부터 시작, 1950년대 번역물까지 소개한다. 원저자, 원작의 제목,

24) 에스카르피, 앞의 책, 121면.

25) Pascale Casanova, *The World Republic of Letters*, 1999, translated by Malcolm DeBevoise, Cambridge: Harvard University Press, 2004.

26) 김병철, 『한국 근대번역문학사 연구』, 을유문화사, 1975.

27) 김병철, 앞의 책, 3면.

번역자, 번역된 작품 제목, 발표지면, 발간일을 일목요연하게 도표로 만들어 보여준다. 또한, 주요 작품들, 원문과 번역문을 비교하여 차이점이 갖는 문학적 의미도 서술한다. 특히 부록으로 첨부된 90페이지 분량의 “서양 문학 번역 연표”(1895~1950)는 이후 등장한 문학 연표의 형식에 중요한 전기를 마련하였다.

김병철의 연표로서의 문학사 기술과 에스카르피의 정량분석에 영향을 받은 국문학자들 중에는 이선영과 권영민이 있다. 이들은 모두, 장기간 문학연구사적 사건(작품 및 평론 출간, 연구논문 간행 등)을 꼼꼼하게 기록하며 이를 활용한 정량분석을 시행한 연구자들이었다. 이선영은 1990년부터 2001년까지 10년 넘게 총 7권의 『한국문학논저 유형별 총목록』을 출간하였는데, 이 저서는 1895년부터 1999년까지 국내외에서 발표한 한국 현대문학 관계 평론과 연구논문을 10년 넘게 정리하고 분류한 결과다.²⁸⁾ 평론과 논문의 이름, 서지사항 등을 김병철의 연표와 비슷하게 일목요연하게 정리하였고 이를 문학 총론, 문학사, 시론, 소설론, 희곡·영화론, 아동 문학론, 번역문학론 등의 주제로 분류하였다. 당시에는 획기적으로 컴퓨터를 활용하여 데이터를 입력하고 정리하였다. 눈에 띄는 대목은, 연도별 작가론의 대상자 수를 제시한 부분이다.²⁹⁾ 1910년대 작가 3명에 대한 작가론은 3편, 1920년대에는 작가 81명에 대해 141편의 작가론이 발표된 데이터를 근거로, 이선영은 1920년대에야 비로소 비평의 대상으로서의 작가가 성립되었다고 주장한다.³⁰⁾ 더불어 그는 1920년대 비평이 발표된 매체 종류에 관한 통계를 작성했는데, 잡지, 신문에 발표된 문학 관계 논문의 수가 전체 90.34%로(잡지: 56%, 신문: 34.43%), 이러한 결과는 비평형성에 있어 저널리즘의 영역이 매우 중요했음을 다시 한번 확인해준다.³¹⁾

이선영은, 『한국문학논저 유형별 총목록』의 결과를 통계적으로 분석한 책을 『한국문학의 사회학』이라는 저서로 출간했다.³²⁾ 서문에서 에스카르피의 문학사회학에 영향을 받아 “작가의 사회의 있어서의 위치”를 살펴봄이 목적이었다고 적고 있다. 책의 전반부는 비평 자체에 관한 정보로 예를 들면, 비평 유형별 연도별 증감, 출판유형별 빈도증감, 가장 활발하게 활동한 비평가 등등을 소개한다. 책의 후반부는, 이 부분이

28) 이선영, 『한국문학논저 유형별 총목록』, 총 7권, 한국문화사, 1990~2001.

29) 이선영, 앞의 책 3권, 711면.

30) 이선영, 앞의 책, 711~712면.

31) 이선영, 앞의 책, 719~722면.

32) 이선영, 『한국문학의 사회학』, 태학사, 1993.

더 중요한데, 한국문학 비평의 중심이 되었던 작가들을 분석한다. 작가론의 대상이 된 상위 200명(전후 시기 포함)의 작가들을 뽑아 등단 시기, 등단 연령, 등단 매체, 출신 지역, 학력, 직업, 사망원인 등등을 분류하여 소개했다. 이를 통해 등단 작가들의 사회적 성격을 살펴볼 수 있다. 1) 식민지시기에 초점을 맞추어 이를 살펴보면, 작가들은 1920~30년대 사이에 집중적으로 등단하였고(82%), 2) 매우 젊었으며(전체 200명 대상, 19~25세: 56.5%) 3) 남성 중심이었고(전체 200명 대상 여성 작가: 7.5%), 4) 잡지와 신문을 통해 등단했다(전체 200명 대상, 잡지: 66%, 신문: 32.2%, 단행본은 3.7%). 또한, 출신 지역 작가 전체로 봤을 때 서울이 제일 많지만 특이하게 1910년대의 경우, 평안남도에서 작가들이 많이 배출되었다(서울보다 인구 100만 명당 작가수가 높다).³³⁾

이선영의 연표가 한국문학 비평과 연구논문에 집중했다면 권영민의 『한국 현대문학 작품연표』는 시와 소설까지 망라한다.³⁴⁾ 이 연표는 그가 한국문학의 시인, 소설가, 비평가를 『한국 근대 문인 대사전』(1990)과 『한국 현대 문인 대사전』(1991)이라는 참고서적으로 정리하면서 그 자료에서 전기적 요소를 빼고 연표형식으로 재정렬하면서 만든 자료다.³⁵⁾ 수필, 희곡, 아동문학, 번역 등등의 정보가 빠져 있고 판본에 관한 설명이 충분치 않으며 빠진 자료 및 오기된 자료도 많다. 그렇지만 작가나 작품생산의 역사적 맥락을 이해하는데 중요한 참고서적이고 한국문학의 계량적 연구에 있어 빠지지 않는 중요한 자료라고 할 수 있다.

특히 그가 간행한 『한국 현대소설 100년』(1995)은 1894년부터 1994년까지 100년간 한국 내에서 신문, 잡지, 단행본으로 간행된 소설 및 소설집을 총정리한 연표와 작가별로 분류한 작품목록을 제공한다.³⁶⁾ 이러한 연표와 목록과 더불어 한국작가들에 관한 계량적 지표와 소설 생산의 그래프를 제시한다. 앞서 이선영은 비평 대상으로서의 작가를 구하여 고빈도 200명을 찾아 그 연도별 증감을 살펴보았다면, 권영민은 시, 소설, 수필, 평론 등 5회 이상 발표한 문인을 일반적 의미의 문필가(1회 발표)와는 다른 문인이라고 보고, 이 전업 작가 2388명의 장르별 분포, 해방 이전 이후의 활동상

33) 이선영, 앞의 1993 책, 141면.

34) 권영민 편, 『한국 현대문학 작품 연표』, 전 2권, 서울대학교 출판부, 1998.

35) 권영민 편, 『한국 근대 문인 대사전』, 아세아문화사, 1990; 『한국 현대 문인 대사전』, 아세아문화사, 1991.

36) 권영민, 『한국 현대소설 100년』, 전 2권, 동아출판사, 1995.

향, 연도별 등단 문인 수, 등단방식 등등을 조사하였다. 이러한 작가 관련 정보와 함께, 권영민은 5년 단위 소설 생산량을 그래프로 보여주고 있다. 이를 식민지 시기에 초점을 맞추어 다시 인용하면 아래와 같다.

〈표 1〉 식민지 시기 5년 단위 소설 생산량³⁷⁾

연도	소설량	연도	소설량
1910	111	1930	575
1915	220	1935	930
1920	103	1940	1200
1925	348	1945	281

위의 표를 보면 각 시기는 이전 5년 동안보다 대체로 증가추세를 보인다. 그러나 두 시기에서 마이너스 성장을 하고 있음을 확인할 수 있다. 물론 모든 자료가 망라되어 자료의 신빙성을 확인할 수는 없지만, 그 쇠락의 폭이 큰 주목할 부분이다. 1915~1920 기간에서 -113%, 1940~1945 기간이 -327%를 기록한다. 역사적으로는 1차대전 직후 국내에 대규모 독립운동과 학살이 있던 시기와 2차대전 직후 일본제국의 총동원 시기와 겹친다. 물론 이 도표가 없더라도 충분히 추론 가능한 결과이지만, 이 도표를 보면 ‘얼마나’ 감소했는지 혹은 성장했는지를 파악할 수 있다. 예를 들어 5년간 소설 편수는 1940년에 최대정점을 찍었지만, 증가율은 1920~1925년 사이가 최대였으며 그 증가율(70%)은 1935~1940년(23%)의 세 배 정도였다는 것과 같은 방식으로 지표를 활용할 수 있다.

연표와 도표에 근거한 1980년대와 1990년대 초중반까지의 문학사회학이 미처 반영하지 못한 요소가 두 가지 있다. 그 중 첫 번째는, 문학주체 사이의 관계론으로, 문학사 자료의 정제형식으로서 연표나 도표가 갖는 속성형 성격 때문에 만들어진 것이다. 연표와 목록, 특히 이선영이 권영민이 정리했던 작가 관련 기록은, 각 시기 한국문학에서 활발하게 활동했던 작가가 누구였는지, 그 집단의 특성을 실증적으로 규명한 기록이다. 역사적 자료를 수집하고 이를 의미 있는 항목으로 분류하고 도표라는 양식으로 형식화했다. 그런데 이 분류의 항목은 고정되어 있어서 이에 의해 분류된

37) 권영민, 앞의 책, 46면, “〈도표 5〉 연도별 소설작품 및 작품집 발표 분포” 재인용.

성격 역시 변화하지 않는다. 작가 관련 항목—예를 들면, 등단 연도, 성별, 나이—, 작품 관련 항목—발표연도, 장르 등등—을 묶으면 좀 더 의미 있는 해석을 할 수 있지만(예를 들면, 1920년대 여성이 발표한 소설은 식민지 시기 전체 소설의 몇 % 등등), 사회학이라는 학문이 개인보다는 집단에 관한 학문이고, 작가라는 개인이 아니라 집단, 특히 작가집단 간의 상호작용(그들이 누구와 교류했는데, 어떻게 관계를 맺었는지, 어떠한 문학적 영향을 미쳤는지)에 초점을 맞춘다면 그 상호작용의 성격은 이러한 속성형 데이터로는 잘 드러낼 수 없다. 게다가 이 속성형 분류를 확장하면 작가에 관한 문학사회학적 현상을 쉽게 물화시켜버릴 가능성이 있다. 예를 들어 어떻게 첫 근대 동인인 창조가 만들어졌는가에 대한 질문에 간단히, 그들은 평안도 출신이었고, 젊었고(그래서 문학적 패기가 있었고), 남성 중심이었기 때문이라고 답할 수도 있는 것이다. 따라서 나중에 다시 설명하겠지만, 문학사회학의 대상으로서의 작가를 속성론이 아니라 관계론으로 다시 이해할 필요가 있다.

IV. 김영민과 매체 중심의 문학사회학

1980년대와 1990년대 초중반의 문학사회학이 반영하지 못한 영역 중 하나는, 작품 생산에 개입한 신문이나 잡지 미디어의 역할이다. 이 기능은 1997년에 출간된 김영민의 『한국근대소설사』에 의해 명확하게 드러나게 되었다.³⁸⁾ 백철의 『조선신문학사조사』 등에서 언급한 이래, 문학연구사에서는 신문 및 잡지 매체를 서구문예사조의 수입 및 유통 통로로 이해하였다.³⁹⁾ 그러나 김영민은 매체가 매개한 문학 및 사회 지식과 더불어 이를 통해 새롭게 등장한 문체와 문학 형식에 주목하였다. 특히 개화기 신문이 유통하던 세계에 대한 백과사전적 지식을, 국내의 독자들에게 쉽게 전달하고 의견을 덧붙이던 수사적 방식과 이에 허구적 양식을 덧붙여 작가/편집자의 의견을 피력하는 양식으로 전환되는 과정을 발견하고, 전자를 “서사적 논설” 후자를 “논설적 서사”라고 불렀다.⁴⁰⁾ 이 단편서사물은 개화기 신문의 한 칼럼이었던 “소설”란에 등장

38) 김영민, 『한국근대소설사』, 숲, 1997.

39) 백철, 『조선신문학사조사』, 전 2권, 수선사, 1947; 백양당, 1949.

40) 김영민, 앞의 책, 29~31, 51~56면.

하기도 했었는데, 이 서사 형식이 신소설 장르로 확장하는 과정의 고찰은, 이광수의 『무정』과 같은 서구식 장편 소설의 도입 이전에, 매체를 활용한 문체실험이 있었음을 증명한다는 측면에서 문학연구사에 큰 획을 그었다. 그의 매체론과 문체론을 결합한 소설양식론은, 그러한 문체 등장의 사회적 성격에 관한 설명이 미흡하다는 비판도 있었지만, 이후의 연구자들에게 큰 영향을 미쳤다.⁴¹⁾ 특히 “한국근대소설사에서 매체는 언제나 작가들의 글쓰기 전략과 방식을 결정한 가장 중요한 변수였다는” 주장, 즉 작가가 매체를 만든 것이 아니라 매체가 작가를 만들었다는 전복적 사고는 신문과 잡지 매체론을 연구한 후학들이 크게 빛지고 있는 부분이다.⁴²⁾

V. 정량분석, 매체연구, 검열연구

계량적 문학연구와 매체연구라는 두 문학사회학적 흐름은, 그간 독자적 영역에서 발전하고 있던 검열연구와 합류하였다. 검열연구는 그간 정진석과 같은 언론학자들,⁴³⁾ 정근식과 같은 사회학자들,⁴⁴⁾ 한만수, 한기형, 박현호, 이해령, 최경희와 같은 문학연구자들⁴⁵⁾ 1990년대와 2000년대를 거쳐 개별적 혹은 검열연구회⁴⁶⁾와 같은

41) 직접적으로는 정선태, 『개화기 신문논설의 서사수용양상』, 소명출판, 1999; 한기형, 『한국근대소설사의 시각』, 소명출판, 1999; 2000년대 초반 상허학보를 통해 등장한 잡지 연구 등이 있다. 영미문학 연구의 대상은 주로 단행본이고 신문잡지 매체는 그에 비하면 주변적이라 할 수 있다. 그러나 미국의 리틀 매거진("little magazines") 연구는 비상업적, 지식인 중심의 잡지가 모더니즘이나 할렘 르네상스의 발흥에 끼친 영향 등을 다루었다. 또한 디지털 아카이브의 도입에 따라 분석가 능한 신문잡지 텍스트가 늘어나면서, 이를 정기간행물 연구(periodical studies)의 틀을 통해 이해 하자는 제안이 최근 등장하기도 하였다. Sean Latham and Robert Scholes, "The Rise of Periodical Studies," *PMLA* Vol. 121, No. 2, 2006, pp. 517~531. 문학사의 국지적 영향관계를 다루는 리틀 매거진 연구에 비해, 글로벌 문학장의 주변부에 있는 한국의 경우, 매체 연구는 근대문학의 형성에 끼친 영향을 탐색한다는 점에서 그 의미는 보다 전방위적이라고 할 수 있다.

42) 김영민, 앞의 책, 496면.

43) 정진석, 『극비 조선총독부의 언론검열과 탄압』, 커뮤니케이션북스, 2007.

44) 정근식, 「식민지적 검열의 역사적 기원-1904~1910」, 『사회와 역사』 64, 2003 외 다수.

45) 단행본 중심으로 이들의 저작을 살펴보면, 한만수, 『허용된 불온-식민지 시기 검열과 한국문학』, 소명출판, 2015; 정근식·한기형·이해령·고노 겐스케·고영란, 『검열의 제국』, 푸른역사, 2016; 한기형, 『식민지 문역-검열, 이중출판 시장, 피식민자의 문장』, 성균관대학교 출판부, 2019 등.

협업을 통해 이론 연구의 성과라고 할 수 있다. 이들은, 검열주체와 검열기구 및 제도, 매체에 따른 검열과정 차이와 작가들의 검열 회피방식, 삭제된 텍스트의 복원과 복원된 텍스트를 고려한 해석의 확장과 같은 여러 주제를 살펴보았다. 또한, 이들의 검열연구는 문학 생산의 개입 기제로서 매체에 관한 연구로 확장되었는데, 신문이나 잡지 미디어를 두고 식민주의자들과 피식민지들이 벌인 경합의 구도, 의도치 않은 연계의 방식, 또한 거시적 측면에서, 근대지식의 비균질적 유통망까지도 살펴보았다.

국문학 내 검열 연구를 심화시킨 계기 중 하나는 『조선출판경찰월보』(1928~1938)라는, 조선총독부 경무국 도서관에서 발간한 비밀자료가 발굴되면서라고 할 수 있다. 이전에도 『신문지요람』(1927), 『출판경찰개요』(1929, 1930, 1932), 『조선총독부 금지 단행본 목록』(1941)와 같이, 식민지시기에 제작된 검열 관계 통계문서는 있었다. 그러나 이 자료만큼 방대하게 검열대상 텍스트를 다루고 치밀하게 행정처분을 항목화 하여 오랜 시간 동안 통계로 만든 자료는 드물었다. 『월보』는 1928년 9월부터 1938년 11월까지 123호가 발행되었고 현재 111호가 확인되었다. 이 자료는 크게 세 부분으로 구분할 수 있는데, A) 검열대상 출판물의 종류와 성격에 관한 개요 및 행정처분 통계, B) 주의, 삭제, 차압, 발매금지 등 행정 처분을 받은 텍스트의 목록, 그리고 C) B)의 행정 처분을 받은 문서의 일본어 요약이다.

2003년 『조선출판경찰월보』가 공개된 이후, 여러 학자가 이 식민지 검열기구가 남긴 기록을 정량적으로 분석하였다.⁴⁷⁾ 박헌호와 손성준의 「한국 근대문학 검열연구의 통계적 접근을 위한 시론」도 그중 하나다.⁴⁸⁾ 이 논문이 통계적 검열연구에 이바지한 바를 몇 가지로 살펴보면, 우선 두 저자는 『월보』 통계자료의 누락과 망실을 전월 기록과 누계를 활용하여 보충하고, 『월보』 통계를 다시 재구성하여 자료 자체에 내재했던 계산 착오나 통계적 오류를 바로잡았다. 이를 통해 이후 학자들이 『월보』를 정확하게 사용할 수 있도록 기초를 닦았다. 또한 『월보』 통계에 대한 정량접근과 더불어, 『월보』의 지면증감 추이를 확인하고 식민지 조선의 인구증감 추이 등 보조통

46) 정근식·최경희, 「도서관의 설치와 일제 식민지출판경찰의 체계화, 1926~1929」, 『한국문학연구』 30, 2007, 103~169면; 검열연구회, 『식민지 검열-텍스트, 제도, 실천』, 소명출판, 2011.

47) 『조선출판경찰월보』에 관한 연구는 이상경, 「『조선출판경찰월보』에 나타난 문학작품 검열 양상 연구」, 『한국근대문학연구』 17, 2008, 389~422면 등 참조.

48) 박헌호·손성준, 「한국 근대문학 검열연구의 통계적 접근을 위한 시론『조선출판경찰월보』와 식민지 조선의 구텐베르크 은하계」, 『외국문학연구』 38, 2010, 193~224면.

계를 활용하여 의미 있는 분석지표를 도출하였다. 예를 들면, 특정 인구집단(일본인/조선인)이 당시의 지식과 자본을 “출판이라는 사회적 행위로 연결할 수 있는” 수행능력을 출판력이라는 지표로 제시하였는데, 이는 이후 연구자들에게도 사용되었다.⁴⁹⁾

여기서 한발 더 나아가 두 저자는 계량적 분석을 통해 검열 적용 방식의 이론화까지 시도하였다. 이를 위해 우선, ‘내지’ 일본과 식민지 조선의 출판력의 차이를, 검열을 받기 위해 납본한 간행물의 발행처를 비교하여 드러냈다. 출판법에 규제받은 국내 간행물과 출판규칙에 의해 규제받은 국외 유수입물을 비교하면, 국내 간행물은 22% 내외(잡지:10%; 단행본: 11%; 기타: 1%)인데 반대 국외 유수입물은 79%나(잡지: 63%; 단행본: 10%; 기타: 6%)된다.⁵⁰⁾ 차압 통계를 살펴보아도 일본(45%)이나 외국 발행(49%) 간행물이 국내 발행물(6%)보다 압도적으로 많았음을 확인할 수 있다.⁵¹⁾ 두 저자는, 이러한 통계들은 식민지 검열관들이 조선에 유입되던 국외출판물을 전력으로 막고 있었음을 보여주는 것이라고 주장하며, 이러한 유수입 통계를 좀 더 큰 틀에서, 제국과 식민지 관계가 놓인 지식장을 존재론적으로 파악하는 근거로 삼았다. 내지보다 강화된 검열제도인 신문지법과 출판법은 조선인에게, 출판규칙과 신문지규칙은 일본인 및 외국인에게 적용되었다. 저자들은, 이러한 ‘법역’의 차이에 의해 식민지 조선은 “‘근대 텍스트의 바다’에 ‘섬’으로 존재했다고 역설하였다.⁵²⁾ ‘문역’과 함께 쌍을 이루는 개념으로 한기형이 정초한 ‘법역’은 식민지 검열이 매개한 조선의 지식생산 및 유통의 장을 이해하는 중요한 개념이다.⁵³⁾ 조금 뒤에 더 설명하겠다.

박현호와 손성준이 위와 같은 검열자가 구축한 통계시스템을 정량적으로 재구성하여, 식민지 조선의 지식유통을 파악하였다면, 이혜령은 같은 『월보』를 분석하여, 검열을 통해 드러난 제국 일본의 위상에 초점을 맞췄다.⁵⁴⁾ 이혜령은, 검열연구는 출판물의 물질성에 대한 탐구라고 전제한다. 일찍이 최경희가 강경애의 “지하촌”에 나타난

49) 이혜령, 「식민지 검열과 ‘식민지-제국’ 표상-『조선출판경찰월보』의 다섯 가지 통계표가 말해주는 것」, 『대동문화연구』 72, 2010, 496면.

50) 박현호·손성준, 앞의 논문, 215면.

51) 박현호·손성준, 앞의 논문, 203면.

52) 박현호·손성준, 앞의 논문, 205면.

53) 한기형, 「법역과 ‘문역’-제국 내부의 표현력 차이와 출판시장」, 『민족문학사연구』 44, 2010, 309~339면.

54) 이혜령, 앞의 논문, 495면.

불구자 인물군을 검열에 의해 텍스트에 가해진 不具化의 연장선상에서 파악한 것처럼, 이해령은 행정처분을 마치 “출판물에 가한 體刑과 같은” 것으로 이해한다.⁵⁵⁾ 행정처분과 사법처분은 검열 결과의 표상방식이다. 이해령이 본 핵심적인 문제는, 출판물의 물질성을 어떻게 정신적 사상적 통제를 겨냥한 검열표준을 기준으로 “계량적인 위계화”를 할 것인가 하는 것이었다.⁵⁶⁾ 이해령은 이 위계화를, 출판책임자에 의해 출판물이 제작되는 과정과 출판물이 독자에게 전달되는 과정 중, 두 번째에 초점을 맞추어 살펴보고자 하였다. 그러나 정치한 질문의 대상과 목적에도 불구하고 논문은 소기의 목적을 달성했다고 보기는 어렵다. 이 계량적 위계화와 관련하여 검열 연구자들에게 흥미로운 부분은, 『월보』의 체계 중 행정 처분을 받은 텍스트의 목록과 일본어 요약의 연결 부분일 것으로 보이는데, 행정 처분을 받은 모든 텍스트가 일본어 요약문으로 제시되는 것은 아니기 때문이다. 이해령은 이것을 샘플링의 문제로 접근하여 어떠한 선택 과정에 의해 목록 일부에 관해 요약 제시문이 제시되었는지를 질문하고 있지만, 이를 비교할 수 있는 별도의 데이터가 아직 만들어지지 않아 질문을 던지는 것에 그치고 있다.⁵⁷⁾

논문에서 이해령은 앞서 언급한 두 번째 과정, 즉 출판물이 독자에게 전달되는 과정의 검열에서 검열관 통계의 특정 항목을 주목한다. 차압 통계 중 빼라 항목이다. 격문이나 선언서 등의 빼라는 검열을 거치지 않고 출판자-독자가 직접 연결되는 비밀출판의 한 종류다. 합법적 출판이 아니므로 차압을 통해서만 통제가 가능하다. 이 차압빼라 통계를 시계열로 분석한 결과에 따르면, 『월보』의 초기부터 전체 차압 건수와 빼라 차압 건수가 비례하여 증가하다가 1929~1930년 사이에 정점에 이른다. 발행처를 살펴보면 외국, 특히 일본에서 만들어져 유입된 빼라가 압도적으로 많고(69%) 그 중 일본에서 조선어로 제작되어 수입된 빼라가 많다.⁵⁸⁾ 빼라는 여러 사회 운동과 불규칙적으로 결합하기도 하고 검열 때문에 내용이 드러나지 않더라도 그 활동이 신문 보도를 통해 이차로 퍼진다는 점에서 문제적이다. 이러한 빼라의 “불투명한 신원과 무정형적 동태”는 식민주의자들에게는 골칫거리였을 것이며, 한편으로 빼라

55) Kyeong-Hee Choi, “Impaired Body as Colonial Trope: Kang Kyŏng'ae's ‘Underground Village’”, *Pubic Culture* Vol. 13, No. 3, 2001, pp. 431~458; 이해령, 앞의 논문, 493면.

56) 이해령, 앞의 논문, 494면.

57) 이해령, 앞의 논문, 521면.

58) 이해령, 앞의 논문, 513면.

는 존재 자체만으로 식민지인 누군가에게는 “제국 일본의 이질적 외부”라는 시공간을 상상하게 하는 기제로 작동했을 것이라고 이해령은 뼈라의 의의를 설명한다.⁵⁹⁾

박헌호, 손성준, 이해령의 작업은 『출판경찰일보』가 제시한 통계를 재구성, 해석하며 검열연구를 한 발 진전시킨 의의가 있다. 그렇지만 부분적, 파편적이라는 한계도 있다. 이러한 검열 내 정량분석을 좀더 거시적인 출판과 표현의 영역으로 확대하여 식민지 지식장의 운동을 개념화한 사람이 바로 한기형이다. 한기형의 초기 관심은, 김영민의 매체 중심의 문학사회학의 연장선상에서, 이를 개화기/식민지 출판문화를 통해 이해하는 것이었다. 그는 이를 검열이 촉발한 제국과 식민지 사이의 출판력과 표현력의 차이로 발전시켰다. 특히, 그는 『월보』와 함께, 『조선출판경찰연보』와 『금지단행본목록』과 같은 검열통계자료를 활용하여 식민지 조선 내 이중출판시장의 새로운 의미를 발견하였다.⁶⁰⁾ 천정환의 『근대의 책읽기』와 같은 책에도 언급되어 있지만, 개화기와 식민지 시기 조선의 출판시장에는 신문학의 유입에도 고전소설, 딱지본, 경서류, 족보 등을 읽는 전통적 독자층은 두터웠다.⁶¹⁾ 한기형은 이들이 읽은 독서물을 여러 가지 표를 이를 더 세밀하게 파악하였는데 1928~1938년 사이 출간된 단행본 통계에 한정하여 보면, “구소설, 신소설, 시가 등 문학 관련 비중이 16.1%, 족보, 유고, 문집 등 전통문화와 관련된 한문 자료의 비중이 27.27%”이다.⁶²⁾ 이 구소설류와 족보류가 전체 단행본 출간의 43.37%를 차지했다. 천정환은 전통서적류 발행을 전통적 독자층이 근대적 대중독자층으로 이행하는 측면에서 중시한 바 있는데,⁶³⁾ 한기형은 더 나아가, 이를 조선 내 출판이 비교우위를 지니는 지점이라고 파악하고 여기서 식민지 토착성이라는 개념을 도출한다. 즉, 족보의 생산자와 구소설의 소비자는 비슷한 멘탈리티의 소유자이며, “과거와 미래를 연결하려는 낭만적 상상력을 통해 그들이 살고 있는 시대의 시간성을 자기 방식으로 해체하고 재규정”하는 이들이라고 주장하였다.⁶⁴⁾

59) 이해령, 앞의 논문, 517~518면.

60) 한기형, 「이중출판시장과 식민지 검열·토착성이라는 문제의식의 제기」, 『민족문학사연구』 57, 2015, 117~154면.

61) 천정환, 『근대의 책읽기-독자의 탄생과 한국 근대문학』, 푸른역사, 2003, 52~58면.

62) 한기형, 앞의 2015 논문, 135면.

63) 천정환, 앞의 책.

64) 한기형, 앞의 2015 논문, 149면.

한기형은 이와 같은, 매체 대상의 계량적 매체 문학사회학을 검열연구와 연결하여 에스카르피가 제시한 문학의 고기압-저기압 지대와 비슷한, 식민지-피식민지 사이의 지식 “분배회로”를 찾아 개념화하려고 시도하였다. 위에 언급한 이중출판시장은, 매체에 적용되는 다른 검열제도—사전검열을 진행한 출판법(잡지 및 단행본 대상)과 사후검열을 진행한 신문지법(신문 및 특별히 허가된 몇 종의 잡지)—⁶⁵⁾와 더불어 대량의 이수입 출판물과 국내 발행 출간물이 경쟁한 결과다. 이러한 이중 출판시장의 존재를 통해 한기형은, 제국-식민지 사이의 커다란 지식 유통장 내 비균질적 영역을 법역과 문역이라는 개념을 만들어 설명하였다. 알튀세르의 국가장치(state apparatus) 개념을 제국과 식민지라는 지역성 위에 올려 놓은 것으로 보이는 법역은, “법과 규칙, 행정행위 등 검열수단이 특정 지역과 그 인구집단에 미치는 효력범위”를 말하며 문역은, 한 나라 안에서 “법률의 지역 간 차이가 만들어내는 표현력의 상대적 편차”를 말한다.⁶⁶⁾ 식민지 조선에서 합법적으로 발화할 수 있는 내용과 주제의 총량은 제국 일본보다 현저히 떨어졌다. 이로 인해, 사회주의 서적이라도 일본에서는 발행되고 조선에서는 금지되어 일본 발행 사회주의 서적이 조선으로 수입되는 양상도 나타났다. 일본 출판업자는 재고떨이라고 여겼던 식민지로의 서적 수출은, 조선으로서는 “오랜 시간 억압되어왔던 식민지 문역의 한계를 해소할 기회”가 된 것이다.⁶⁷⁾ 문역이 법역의 종속변수인지, 법역이 알튀세르의 국가기구라면 문역은 이에 대항하는 시민 영역의 반국가기구인지, 검열과 출판제도 이외에 이러한 대립적 영역을 구성하는 요소는 무엇인지 아직 해소되지 않은 의문들이 남아있다. 그렇지만 적어도 이 법역과 문역 개념은, 에스카르피가 보지 못한 식민지 내외의 서적 분배회로와 이 회로에 미치는 검열의 영향을 파악할 연구들을 제시한다.

Ⅶ. 근현대 한국문학의 멀리서 읽기

한편, 2010년대에는 이선영과 권영민 등이 기초를 닦은 계량적 문학사회학을 잡지

65) 한기형, 『식민지 문역-검열, 이중출판 시장, 피식민자의 문장』, 성균관대학교 출판부, 2019, 168면.

66) 한기형, 앞의 책, 68~69면.

67) 한기형, 앞의 책, 86면.

에 특화한 공동연구가 등장하였다. ‘한국근대잡지 소재 문학텍스트’라는 작업이 그것인데, 2010년부터 2013년까지 고려대 한국학 연구소 중심으로 진행한 한국연구재단 인문사회연구 토대기초연구의 결과물이다. 1906년부터 1945년까지 11,285명의 필자가 잡지에 기고한 50,939편의 텍스트를 모아서 장르별로 분류했다.⁶⁸⁾ 아홉 명의 연구원이 꼼꼼하게 모은 자료는 규모도 방대할뿐더러 이전의 오식이나 서지학적 오류로 잡은 것으로 보인다. 결과물은 처음에는 데이터베이스 형태로도 제작되었으나 링크는 현재 사라졌고 현재 네 권의 책으로 남아 있다.⁶⁹⁾ 이 자료의 성과는 「한국근대잡지 소재 문학텍스트에 대한 통계적 분석」이라는 논문에 담겨있다. 식민지시기 작품(시, 소설, 수필, 평론, 희곡) 생산량을 시계열로 분석하였고,⁷⁰⁾ 더불어 시기별 장르 비율, 장르별 작가 비율의 추이도 제시한다. 이러한 커진 규모의 통계는 식민지 문학장의 생산력을 가늠하는 실증적 자료로서 여전히 중요한 의미가 있지만, 앞서 언급한 잡지가 매개한 문학생산관계, 즉 잡지 매체가 매개한 문학생산주체와 작품 사이의 역동적 상호관계를 살펴보기에는 미흡한 점이 있다.

이처럼 작가 수와 작품 수와 같은 단순 통계를 통해 문학사를 이해하는 기존의 연구에서 한발 더 나아가고자 하는 요구는 2000년대 이후 조금씩 보편화 된 디지털 인문학 연구 방법의 성과다. 디지털 인문학은 갑자기 등장한 것은 아니고, 개별 학문 분야(예를 들면, 사회학, 언어학 등)에서 디지털 아카이브를 만들고 컴퓨터를 활용하던 여러 방법이 각 학문적 경계를 넘어 적용되면서 나타난 방법론적 현상이라 할 수 있다.⁷¹⁾ 사회학에서 정보의 이동경로 분석, 회사 내 소집단 관계 파악에 사용되던 네트워크 분석 방법이 문학주체와 매체 사이의 관계를 이해하는데 적용되고 언어학에서 사용하던 말뭉치(코퍼스) 분석은 역사학 분야, 특히 개념사 분야에 적용되어 오늘날과는 다른, 당대의 개념어가 가진 의미망을 분석하는 데에 사용되었다.⁷²⁾ 그리

68) 전도현, 「한국 근대잡지 소재 문학텍스트에 대한 통계적 분석」, 『한국학연구』 47, 2013, 131~159면.

69) <http://modernjournal.org> 이 데이터베이스의 링크는 더 이상 작동하지 않는다. 최동호, 최유찬 외 편저, 『한국 근대잡지 소재 문학텍스트 연구』, 총 4권, 2012~2013.

70) 전도현, 앞의 논문, 137면. 1930년대 후반이 최고용성기(최고점은 1936년, 연간 발행 작품수가 4162편)이고 이후 급감하는 모양을 보인다.

71) 송인재, 「신기한 인문학」, 『세계 디지털 인문학의 현황과 전망』, 커뮤니케이션북스, 2019.

72) 허수, 「어휘연결망을 통해 본 ‘제국’의 의미-‘제국주의’와 ‘제국’을 중심으로」, 『대동문화연구』 87, 2014, 501~561면; 「네트워크 분석을 통해 본 1980년대 ‘민중’-『동아일보』의 용례를 중심으로」, 『개념과 소통』 18, 2016, 53~95면.

고 이러한 디지털 인문학의 방법론은 서론에서 언급한 것과 같이 모레티가 주창한 ‘멀리서 읽기’와 접합하여, 한국에서도 이를 중심으로 여러 방법론적 자장이 형성되었다.

이 멀리서 읽기의 연장선상에서 국문학 분야의 문학사회학적 연구사례는 크게 네 가지로 분류할 수 있다. 1) 텍스트 분석을 포함하지 않는 문학집단이나 군집분석 2) 말뭉치 분석에 기반한 어휘망 분석(네트워크 분석의 어휘/의미론적 적용) 3) 어휘 분석에 딥 러닝적용(AI를 사용한 심화학습)을 적용한 연구 등이 있고 4) 고전문학 분야에서는, GIS(지리공간정보)를 활용하여 인물, 시간, 사건 등의 각각 다른 범주의 다른 스케일의 데이터를 연결하여 새로운 의미를 찾고 있다.

1)과 관련하여 이재연은, 위에서 언급한 “한국 근대잡지 소재 문학텍스트” 데이터베이스에서 데이터를 얻고, 이 데이터베이스에서 빠진 잡지와 주요 신문에 게재된 소설을 직접 조사하여 1917년부터 27년 사이 주요 신문과 잡지에 게재된 소설 목록을 만들었다. 이를 바탕으로 소설과 매체(신문과 잡지) 사이의 관계를 사회연결망 분석으로 파악하였다. 이 네트워크 분석은 이전에 통계 분석에서 드러낸 근대작가의 정태적 성격에 역동성을 부여했다는 의의가 있다. 다시 말해, 이전의 젊은 나이에 등단한 남성 중심의 작가들이라는 일반적 속성 외에, 작가들이 어느 지면을 통해 창작 활동을 했는가에 따라 그 작가 군집과 매체의 성격 및 역할이 다르게 이해할 수 있음을 보였다. 예를 들어, 1920년대의 총독부 기관지였던 『매일신보』는 상업적으로 라이벌 관계에 있던 『조선일보』나 『동아일보』 필자를 적극적으로 영입하여, 작가의 공유가 빈번하지 않던 두 민간지를 매개하는 역할을 하고 있다.⁷³⁾ 또한, 잡지 투고 작가의 군집을 살펴보면, 당시 『동아일보』, 『동광』 등 소수의 친밀한 잡지와 교류하고 있던 이광수보다는 (생존을 위해) 매체를 가리지 않고 활발한 문학 활동을 했던 최서해나 김명순의 문학장 내 영향력이 크게 나왔다.⁷⁴⁾ 또한, 위의 데이터 중 창조, 폐허, 백조의 3대 동인의 문학 활동을 중심으로, 이를 동인 이전-동인 시기-동인 이후의 세 시기로 나누어 보니 보통 남성 중심 동인 집단의 지적 액세서리처럼 여겨졌던 나혜석이나 김명순이, 동인 역사의 전사(prehistory)로 기능함을 살펴볼 수 있었다.⁷⁵⁾ 이러

73) 이재연, 「작가, 매체, 네트워크-1920년대 소설계의 거시적 조망」, 『사이』, 2014, 285~286면.

74) 이재연, 앞의 논문, 288~291면.

75) Jae-Yon Lee, “Before and after the “Age of Literary Coteries: A Diachronic Analysis of Writers”

한 분석결과를, 물론 데이터의 선별기준에 따라 값이 달라지기는 하지만 이것을 고려 하더라도 우리가 아는 문학사적 상식이 가정의 차원에서 이해되어야 함을, 그 자체로는 절대적이지 않음을 보여준다. 우리는 문학사에 관해, 무엇을 근거로 어디까지 말할 수 있는가를 다시 생각해 보아야 할 시간이 된 것이다. 이런 1920년대 작가 네트워크 연구를 필두로, 전봉관, 이원재, 김병준의 1990년대와 2000년대의 주요 문예잡지를 대상으로 한 작가집단 분석이 등장하였다.⁷⁶⁾

그러나 이와 같은 소셜 네트워크 분석은 여러 범주를 동시에 연결하여 다차원적으로 파악하기는 어렵다는 점에서 한계가 있다. 범주가 같은 일원모드나(작가-작가 혹은 매체-매체) 스프레드시트에서 행과 열로 표현할 수 있는 이원모드(작가-매체)와의 관계 정도만 가능하다. 세 개 이상의 범주와 그 사이의 관계를 동시에 파악하는 작업은, 온톨로지라는 방식을 사용하면 가능하다. 온톨로지는 인간이 인식하는 다른 범주와 척도의 지식을 컴퓨터가 이해할 수 있도록 정형화하여 개념과 개념의 관계를 기술하는 연구 분야나 방식을 말한다. 고전문학 분야에서는 이 온톨로지를 사용하여 많은 연구가 등장했다. 이재옥은, 조선 시대 문과 급제자의 기록인 『문과방목』을 대상으로, 문무과 급제자의 재급제자까지 조사하고 이들의 친인척 관계까지 네트워크로 파악하였다.⁷⁷⁾ 김바로는, 여러 가지(“한국역사정보 통합정보시스템”, “역대인물 종합정보시스템” 등의) 이미 구축된 데이터베이스의 개별 정보를 온톨로지로 연결하여 인물, 공간, 사건, 시간 등을 넘나드는 거시적 고전연구가 가능한지를 타진하였다.⁷⁸⁾ 또한, 류인태는 고산 윤선도의 손자인 윤이후가 쓴 『지암일기』를 공동 번역하고 온라인 데이터베이스로 만들면서 그 일기에 담긴 수십 종의 생활사 데이터를 온라인상에 시각화하였다.⁷⁹⁾ 이러한 온톨로지 방식은 아날로그 정보를 연구자가 일일이 항목을 만들고 계열화해야 하는 수고로움이 있지만 일단 데이터베이스가 구축되면 다른 데

Networks in Korea, 1917~1927”, *Korea Journal* Vol. 57. No. 2, 2017, pp. 35~68.

76) 전봉관, 김병준, 이원재, 「문예지를 매개로 한 한국 소설가들의 사회적 지형」, 『현대소설연구』 61, 2016, 169~228면.

77) 이재옥, 「조선시대 문무과 재급제 현황과 재급제자 조사」, 『장서각』 32, 2014, 168~197면; 이재옥, 『조선시대 과거 합격자의 디지털 아카이브와 인적 관계망』, 보고서, 2018.

78) 김바로, 「역사기록의 전자문서 편찬방법 탐구-역사요소를 중심으로」, 『열상고전연구』 51, 2016, 203~223면.

79) 류인태, 「지암일기-데이터로 다시 읽는 조선시대 양반의 생활」, <http://jamdiary.info> (2020년 6월 2일, 링크 정상 작동).

이터베이스와의 연결과 확장을 통해 지식의 축적을 도모할 수 있는 장점이 있다.

네트워크 분석은 이전에 보이지 않았던 문학적 관계를 보여주었다는 점에서 충분한 의의가 있으나, 그 자체로 본격적인 텍스트 분석은 아니다. 연구자들은 이 방식을 텍스트에 사용된 어휘 분석과 해석에 적용. 이렇게 어휘 분석을 가능하게 하려면, 이미 텍스트가 디지털화(pdf가 아닌, txt 형태로) 저장되어 있어야 한다. 이를 자연어 처리하여 형태소의 수준으로 잘라내면, 문서에 등장하는 고빈도어를 파악할 수 있고 또 고빈도 키워드와 주로 등장하는 어휘(共起語)를 파악할 수 있다. 연구자가 살펴보고자 하는 키워드와 공기어의 빈도와 결합 양태를 통해 그 어휘들이 문장에서 어떠한 의미로 사용되었는지 이해 가능한데, 이렇게 전산 분석을 활용한 결과를 바탕으로 다시 텍스트를 읽는다면 기존의 연구가 발견하지 못한 지점을 파악할 수 있다. 이러한 2)의 방식과 관련하여 이재연은, 1920년대의 중요한 월간 종합지 『개벽』을 토픽 모델링으로 분석하였다.⁸⁰⁾ 토픽모델링은, 키워드의 공기어 중, 키워드와 같이 나타날 확률이 높은 단어(의미상으로 유사한 단어)를 계산하여 군집시키는 방식을 말한다. 이 방식은 실제 통계가 아닌 확률에 바탕을 두고 있으므로 분석조건이 달라지면 값이 달라지는 경향이 있기는 하지만(그러나 다른 값을 주고 시도하더라도 결국 비슷한 패턴으로 수렴한다), 한 연구자가 완전히 읽어내지는 못하는 문서 속 전체 주제를 의미론적으로 조망할 수 있는 장점이 있다.

이 토픽모델링을 적용하여 그는 『개벽』의 전체 주제의 지도를 그려 시각화하였다. 이 지도의 중심엔 이미 여러 연구자가 이 잡지의 핵심주제로 파악한 ‘사회’와 ‘개조론’이 위치한다. 한편 문학 관련 주제는, 자체로서는 잡지의 핵심적 위치를 차지하고 있지는 않지만, ‘사회’와 ‘개조론’을 잇는 주요 주제로 네트워크에 연결되어 있음을 확인할 수 있었다. 또한, 낭만주의 문학 작품과 담론에 주로 사용되던 ‘생명’과 사실주의나 좌파문학의 핵심어였던 ‘생활’의 공기어를 토픽모델링으로 파악하여 이 두 주제가 서로의 문맥을 전경화하거나 혹은 후경화하여 의미를 변화시켜나가는 과정을 포착하였다. 한 개념의 변주는 방계 개념들 사이의 의미론적 상호작용 없이는 불가능했던 것이다. 커다란 주제 네트워크 위에서의 지엽적 변화가 어떻게 시간이 지나면서 개념적 다양성 사이에 변화를 일으켜 네트워크 시스템 전반에 영향을 미쳤는가, 스티

80) 이재연, 「키워드와 네트워크-토픽 모델링으로 본 『개벽』의 주제지도 분석」, 『상학학보』 46, 2016, 277~334면.

본 제이 굴드식으로 표현하자면 어떻게 야구 시스템의 전반적 향상이 4할대 타자를 멸종시켰나, 하는 데까지는 나아가지는 못했다.⁸¹⁾ 그러나 전체 주제의 망 속에 잡힌 지엽적 개념의 변주는 신문, 잡지 내 개념 변화의 추이를 정량분석으로 접근하는 하나의 참조를 제공한다.

3) 텍스트에 사용된 어휘의 전산 분석 중 기계학습(machine learning)을 사용한 방식은 최근에 등장한 것이다. 이세돌과 알파고의 바둑대결에서 유명해진 이 방식은, 컴퓨터에 수많은 데이터를 학습시키고 훈련하여 주어진 문제를 해결할 수 있는 최적의 경로와 패턴을 찾아내도록 한다. 이 훈련에 개입한 사람의 지도 정도에 따라 지도 학습이나 비지도 학습으로 나뉘는데, 앞서 설명한 공기어 찾기를 발전시킨 ‘워드투벡터’(word-to-vector, 이후 word2vec)는 데이터 처리에 대한 명확한 방향성을 주지 않은 비지도 학습 계열의 알고리즘이다. 보통의 공기어 찾기 프로그램은 키워드의 앞에 등장한 몇 단어 뒤쪽의 몇 단어를 찾아낸다. 키워드에서 멀리 위치한 공기어는 찾을 수 없다. 그러나 이 알고리즘은 키워드의 공기어를 독립 요소가 아닌 연속된 행렬로 구성하는 워드 임베딩(word embedding) 위에, “하나의 단어[를] 미리 정의된 차원에서 연속형의 값을 갖는 벡터” 값으로 표시한다.⁸²⁾ 정의된 차원은, 문맥을 고려하여 구성된 가상공간을 의미하므로 비슷한 의미의 어휘들은 같은 공간에 위치하게 된다. 단어 사이의 거리가 가까울수록 가중치를 주어 벡터값을 구하는데, 이렇게 키워드와 동시 출현하는 수많은 단어를 컴퓨터가 학습하게 되면 키워드에 물리적으로 가까운 단어뿐만 아니라 멀리 떨어져 있어도 의미상으로 가까운 단어까지 포착할 수 있게 된다.

최지명은, 『개벽』의 텍스트에서 필명이거나 무명씨의 논설 텍스트 164편을 선별하여 기계학습(지도학습)의 방식을 발전시켜 작가 판별 모델을 개발하였다.⁸³⁾ 『개벽』

81) 스티븐 제이 굴드 저, 이명희 역, 『폴하우스』, 사이언스북스, 2002, 115~138면.

82) T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, & J. Dean, “Distributed representations of words and phrases and their compositionality”, *Advances in neural information processing systems*, 2013, pp. 3111~3119.; 정재운, 모경현, 서승완, 김창엽, 김해동, 강필성, 「워드 임베딩과 단어 네트워크 분석을 활용한 비지도학습 기반의 문서 다중 범주 가중치 산출 휴대폰 리뷰 사례를 중심으로」, *Journal of the Korean Institute of Industrial Engineers*, Published Online, 2018, 443면에서 재인용.

83) 최지명, 「기계학습을 이용한 역사 텍스트의 저자판별-1920년대 『개벽』 잡지의 논설 텍스트」, 『언어와 정보』 22(1), 2018, 91~122면.

의 총저자를 모르는 상황에서 저자판별은 쉽지 않으므로, 먼저 저자가 될 확률이 높은 작가들의 집단을 구하였다. 먼저, 다작한 저자가(논문에서는 3편 이상을 쓴 29명) 저자를 확인하고자 하는 신원 미상의 논설을 썼을 것으로 가정하고, 이들의 259편 텍스트를 각 저자의 텍스트로 학습시켰다. 이후 이 결과를 필명이나 작가 미상의 텍스트에 적용하여, 문장의 특성(길이)이나 어휘의 특성(종류와 분포) 등이 특정 저자의 것과 유사한지를 분별하는 방식으로 저자를 판별하였다. 이처럼 기계학습을 통한 어휘분석이나 문체분석은, 앞서 살펴본 기초통계 중심의 연구사례에서 기술적으로 상당히 나아갔다. 물론 전산기술의 급속한 진전 때문에 인문학 연구가 사회과학의 아류로 전락하는 것은 아닌가 하는 염려가 없는 것은 아니지만, 동시에 이러한 기술적 진전은 국문학 내 문학사회학의 새로운 가능성과 지평을 열고 있다 하겠다.

VII. 覆字의 멀리서 읽기 시론

검열연구는, 앞 절에서 언급한 국문학 내 멀리서 읽기의 방법론적 유효성을 타진할 수 있는 분야 중 하나다. 검열된 텍스트를 해석하면서 겪는 어려움은 여러 가지가 있지만, 그중 가장 난해한 것은, OO나 XX처럼 삭제된 단어를 대체하여 표기한 覆字(후세지 ふせじ)의 문맥을 파악하는 일이다. 이전에도 삭제된 텍스트를 복원하는 시도는 있었다. 한만수는 강경애의 단편소설 「소금」에서 검열관이 붓으로 지운 230여 자를 복원하였는데, 활판인쇄에서 생긴 壓痕에 적외선과 자외선을 투과시키고, 또 종이 뒤편에서 빛을 투과하는 등 최첨단 감식기술을 동원하였다.⁸⁴⁾ 그러나 이 경우는, 잉크 뒤에 원문 텍스트가 남아있는 운이 좋은 상황이다. 단어나 구절이 저자나 편집자에 의해 이미 복자로 대체된 경우, 이러한 감식방법을 쓰더라도 그 원문을 밝혀 내기는 어렵다. 따라서 우리는, 원문이 무엇인지 모르는 원문을 복원하는 방식이 아닌, 통계를 사용하여 삭제된 텍스트의 맥락을 짐작하는 추론적인 방식을 택하기로 하였다. 복자 주변에 등장한 단어의 빈도와 종류를 구하여 삭제된 단어의 의미를

84) 한만수, 「강경애 '소금'의 복자 복원과 복한 '복원본'의 비교」, 김인환 외 편, 『강경애, 시대와 문학-강경애 탄생 100주년 기념 남북한 공동논문집』, 랜덤하우스코리아, 2006.

유추하는 방식, 즉 복자를 키워드처럼 처리하여 복자의 공기어들을 추출하고 이 공기어들 사이의 관계나 유사성 혹은 패턴을 찾는 방식으로 삭제된 부분의 의미를 추론해 보고자 한다. 이 연구는 현재 진행 중이며, 논문 작성시까지 얻은 관찰의 몇 부분만을 소개하고자 한다. 아직 성근 관찰 결과에 대해 미리 양해를 구한다.

이를 위해 먼저 식민지 시기 잡지인 『별건곤』(1926~1934)과 『삼천리』(1929~1942)를 대상으로, 복자가 어떻게 사용되었는지를 살펴보았다. 두 잡지의 발행 기간이 『조선출판경찰월보』과 겹치는 시기가 있어 향후 『월보』에 관한 기존의 통계연구를 참조할 수 있을 것이라는 기대가 있었고, 두 잡지 모두, 『개벽』과 같은 심각하고 진지한 읽을거리라기보다는 대중취향의 가벼운 읽을거리이기 때문에 이곳에서 복자가 발견된다면 더 다양한 맥락을 담고 있을 것이라 가정했다. 또한, 연구 결과에 따라, 각 잡지의 미디어 전략에 따른 복자의 의미화 방향의 차이를 구할 수도 있다고 예상했다.

텍스트는 국사편찬위원회에서 디지털화한 원문자료와 서지정보를 사용하였다. 우선 이 디지털 텍스트에서 어떤 기호들이 복자로 사용되고 있는지 확인하였다. 가능한 다양한 기호들을 추출하기 위해 간단한 텍스트 전처리 과정을 거쳤다. 문장을 공백 기준의 어절(token) 단위로 분리한 후, 단어와 기호(punctuation)가 같이 나타난 어절만을 남겼다. 기호 중 구두점으로 사용되는 마침표, 느낌표, 물음표, 쉼표 등과 일본어에서 숫자 0으로 쓰이는 ‘〇’도 제외하였다. 마지막으로 숫자와 옛한글, 한글, 한자 문자들을 모두 삭제하여 빈도를 측정하였다. 다음 <표 2>는 말뭉치에 사용된 복자의 종류와 연속으로 사용된 길이에 따른 빈도를 나타낸다. ‘X’라는 기호가 두 글자 연속으로 사용된 경우는 1,413번인데, 이는 두 음절이 ‘XX’로 표기된 횟수를 의미한다. 이 검색의 결과 총 13종의 기호가 복자후보 기호로 등장하였다.

〈표 2〉 복자후보 기호의 종류와 빈도

기호	한 글자	두 글자	세 글자	네 글자	다섯 글자
X	1,529	1,413	214	106	33
x	24	40	4	1	-
X	4	2	-	-	-
X	2	-	-	-	-
x	2	-	-	-	-

*	1,203	296	128	77	45
○	24	11	62		-
O	685	343	76	17	11
o	432	63	-	-	-
o	301	20	32	-	-
■	178	32	11	2	3
□	10	1	-	1	-
△	-	8	1	-	-

위에서 “복자”가 아닌 “복자후보 기호”로 언급한 이유는, 두 가지 의미에서다. 아직 프린트본과 디지털본의 대조를 거치지 않았기 때문에 위의 기호가 어떤 이유에서든 디지털 텍스트에서만 사용되었을 가능성을 열어 두어야 하고, 한편 “00번지”의 예처럼, 고유명사나 숫자 등의 생략을 표시하기 위해 사용된 기호가 있기 때문에, 위의 기호에서 내적 혹은 외적 검열에 의해 발생한 복자를 구별해야 한다. 어떻게 구별할 수 있을지 좀 더 고민이 필요한 대목이다. 여튼, 가장 많이 사용된 기호는 대문자 ‘X’였으며, 소문자, 로마자, 곱하기 기호 등으로 유사하게 표현되었다. 그다음으로는 ‘*’, ‘O’의 순으로 많이 나타났다.

다음 『별건곤』과 『삼천리』를 묶은 전체 기사를 대상으로, 복자후보 기호가 사용된 글의 종류를 살펴보았다. <표 3>은 국사편찬위원회의 원문에 표기된 글의 분류에 따라 기호가 포함된 글이 해당 분류에서 차지하는 비율을 나타낸다. 편수로는 논설과 문예 기타, 소식의 순으로 많이 나타났으나, 비율 순으로는 세태비평, 회고·수기, 논설의 순으로 기호가 많이 포함되어 있었다. 15% 이상, 상대적으로 기호가 많이 등장한 장르를 파랗게 처리하여 『별건곤』과 『삼천리』를 비교하여 보니, 『별건곤』에서는 세태비평, 기행문, 회고·수기 순으로, 『삼천리』에서는 논설, 회고·수기 순이었다. 이 차이가 각 잡지의 미디어 전략과 어떤 관계가 있을까? 한편, 소설에서 기호가 사용된 비율은 12에서 14% 정도로, 위에 언급한 장르보다 상대적으로 낮았다.

〈표 3〉 분류별 복자포함 기사의 비율

전체

분류	복자포함/전체	비율	분류	복자포함/전체	비율
기행문	25/188	13.30%	세태비평	29/153	18.95%
논설	172/1010	17.03%	소설	38/293	12.97%
대담·좌담	36/262	13.74%	소식	108/950	11.37%
만화	1/39	2.56%	시	9/509	1.77%
문예기타	114/945	12.06%	잡저	76/1061	7.16%
문예평론	25/177	14.12%	학술	0/2	0.00%
사고·편집후기	32/566	5.65%	회고·수기	91/534	17.04%
설문	23/233	9.87%	희곡·시나리오	9/58	15.52%

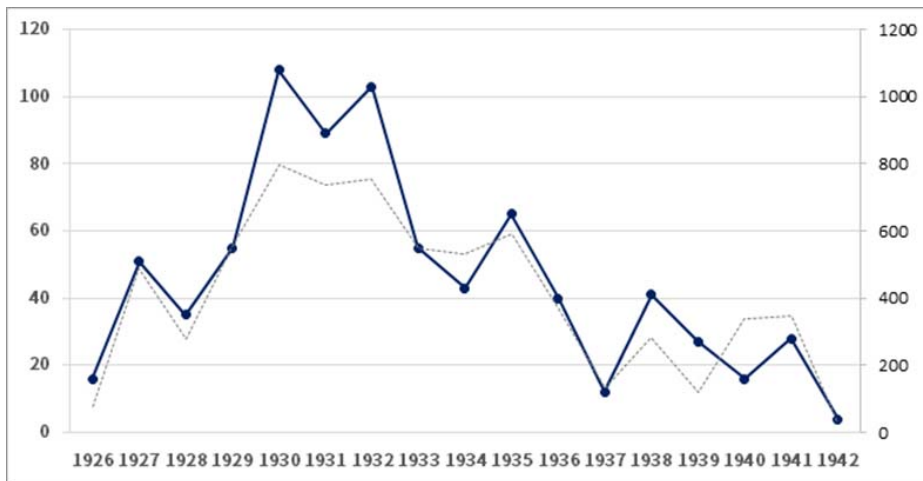
별건곤

분류	복자포함/전체	비율	분류	복자포함/전체	비율
기행문	7/35	20.0%	세태비평	26/122	21.3%
논설	46/346	13.3%	소설	15/102	14.7%
대담·좌담	5/42	11.9%	소식	34/273	12.5%
만화	1/38	2.6%	시	2/109	1.8%
문예기타	58/509	11.4%	잡저	51/589	8.7%
문예평론	2/27	7.4%	학술	0/0	-
사고·편집후기	25/321	7.8%	회고·수기	35/203	17.2%
설문	8/111	7.2%	희곡·시나리오	4/20	20.0%

삼천리

분류	복자포함/전체	비율	분류	복자포함/전체	비율
기행문	18/153	11.76%	세태비평	3/31	9.68%
논설	126/664	18.98%	소설	23/191	12.04%
대담·좌담	31/220	14.09%	소식	74/677	10.93%
만화	0/1	0.00%	시	7/400	1.75%
문예기타	56/436	12.84%	잡저	25/472	5.30%
문예평론	23/150	15.33%	학술	0/2	0.00%
사고·편집후기	7/245	2.86%	회고·수기	56/331	16.92%
설문	15/122	12.30%	희곡·시나리오	5/38	13.16%

다음으로, 복자 사용의 추이를 연도별로 살펴보았다(<그림 1>). 복자가 포함된 기사가 많이 나오던 때는 두 시기로, 1929~1933년과 1937~1939년 사이이다. 이 시기, 전체 기사 건수가 늘어나지만(회색 점선), 이 전체 기사의 발행 추이보다 복자후보 기호가 사용된 기사(파란 실선)가 더 증가하였다. 반면 그 반대인 경우도 있었는데, 1939년부터 1941년 사이이다. 역사적 문학사적 사건의 전개와 기호 포함 잡지 기사의 증감과의 관계를 앞으로 살펴볼 필요가 있다.



〈그림 1〉 복자 후보기호가 나타난 기사의 연도별 빈도 (점선은 연도별 전체 기사의 수)

다음으로는 복자가 사용된 맥락을 살펴보기 위해 복자와 함께 출현한 공기어들을 살펴보았다. 자연어처리 분야에서 단어의 맥락이나 유사어를 파악하는 방법으로 사용되는 것 중 하나가 앞 절에서 간단히 설명한, 단어 임베딩(word embedding)이다. 임베딩은 단어를 수치의 나열인 벡터로 표현하는 것으로, 단어를 벡터로 표현하게 되면 다른 단어와의 유사도 측정이 가능해진다. 단어 벡터화의 가장 기본적인 방법은 문헌-용어 행렬(Document-Term Matrix; DTM)을 구축하는 것인데, 특정 문헌에 나온 단어의 빈도를 행렬화하는 것이다. 다음 <그림 2>에서와 같이 문헌 d_1 에 단어 t_1 이 몇 번 출현했는지 빈도를 채워 넣는 방법이다. 이런 과정을 거치면 음영으로 표시한 부분과 같이 단어 t_1 에 대한 단어벡터 $\{1, 1, 0, \dots, 1\}$ 를 얻을 수 있다. $t_2, t_3, \dots, t_m$ 에 대해서도 동일한 과정을 거쳐 단어의 벡터를 얻게 되면, 행렬계산을

통해 단어의 공기빈도를 손쉽게 얻거나, 벡터 간의 유사도 측정을 통해 단어 간의 거리를 측정할 수 있다.

Document						x	Term						=	Term					
	d_1	d_2	d_3	...	d_m			t_1	t_2	t_3	...	t_m			t_1	t_2	t_3	...	t_m
t_1	1	1	0		1		d_1	1	0	0		1		t_1		7	3		0
t_2	0	1	0		0		d_2	1	1	1		1		t_2 <td></td> <td></td> <td></td> <td></td> <td>1</td>					1
t_3	0	1	1		0		d_3	0	0	1		0		t_3 <td></td> <td></td> <td></td> <td></td> <td>5</td>					5
...							...							...					
t_m	1	1	1		1		d_m	1	0	0		1		t_m <td></td> <td></td> <td></td> <td></td> <td></td>					
MxN							NxM							MxM					

<그림 2> 문헌-용어행렬 곱을 사용한 공기어빈도 측정

단어가 어떤 문헌에 출현했는지를 표현한 용어-문헌행렬(<그림 2>의 좌측)과 문헌에 어떤 단어가 나타났는지를 표현한 문헌-용어행렬을 행렬곱하게 되면 <그림 2>의 우측처럼 용어에 관한 $M \times M$ 크기의 정방대칭행렬(squared symmetric matrix)을 얻을 수 있고, 행렬곱을 통해 얻은 음영 부분의 값은 용어들의 동시출현 빈도가 된다. <그림 2>에서는 t_1 과 t_2 의 값인 7이 두 단어의 동시출현빈도가 되는 것이다.⁸⁵⁾

본 연구에서는 복자와 공기한 단어의 빈도를 측정하기 위해 자주 등장하는 복자후보 기호 ‘XX’, ‘XXX’, ‘**’, ‘***’, ‘OO’, ‘OOO’의 앞뒤로 출현한 다섯 단어씩을 추출하여 동시출현 단어들의 빈도를 측정하였다. 다음 <표 4>는 주요 복자와 동시에 자주 나타난 상위 20단어이다.

85) 문헌-용어 행렬을 사용해서 동시 출현 빈도를 계산하는 방법은 행렬의 대부분이 0으로 채워지기 때문에 희소행렬(sparse matrix)이 될 가능성이 높아 계산의 효율성이 떨어진다는 단점이 있다. 따라서 동시출현빈도를 계산할 때에는 분석단위가 되는 텍스트(문장이나 문헌)에 출현한 동시출현 단어 쌍을 리스트화하고 그 빈도를 누적으로 계산하는 방법을 주로 사용한다. 그러나 단어의 벡터 유사도 측정에 유리하고 대량의 텍스트가 아닐 경우에는 단어 빈도 계산이나 행렬계산이 빨라지며, 단어에 tf-idf 가중치를 부여하는 등 수치의 정규화가 가능하다는 장점이 있다.

〈표 4〉 복자후보별 동시출현단어 상위 20단어 (복자후보 기준 다섯 단어 기준)

순위	XX	XXX	**	***	OO	OOO
1	사람	사람	사람	사람	사람	내용
2	우리	생각	日本	日本	부대	사람
3	朝鮮	사제	우리	관계	은행	탄식
4	주의	우리	생각	까닭	운동	삭제
5	운동	조선	관계	內地	소리	선생
6	조선	교주	까닭	塩原局長	우리	야단
7	재벌	마누라	모양	조직체	친구	無能庸才
8	생각	문제	생명	東京	학교	결박
9	계급	관계	설치	거절	생각	까닭
10	문제	사실	정치	널비	사실	당시
11	푸로레타리아	우리들	지불	다방	가슴	매부
12	부인	친척	支持	모양	목사	법석
13	학교	구타	朝鮮心	문제	문제	사랑
14	노동자	마음	朝鮮質	바람	부인	상점
15	정치	병원	産業	세계	일보	양복
16	사건	생활	能力	세월	결혼	우리
17	신문	얼굴	문제	실수	만원	우리집
18	까닭	예수교	바람	요구	신문	유감
19	남편	제국주의	시작	정치	아래	전당포
20	농민	某新聞社	쌍방	혐의	연애	처녀

위의 공기어는, 예상했음과 의문, 그리고 미진함과 놀라움 사이에 놓인 어떤 것으로 연구자에게 다가온다. ‘계급’, ‘푸로레타리아’, ‘노동자’, ‘정치’, ‘제국주의’ 등 그간 여러 연구를 통해 식민지시기 불온함의 표상으로 연구된 단어들이 등장했음은 예상한 바와 같다. 그런데 한편으로 왜 이 불온한 단어들은 남고 주변의 다른 단어들이 XX, OO 등으로 처리되었을까? 이 기호들 안에는 위의 불온한 단어들보다 더 불온한 단어가 들어있었던 것일까? 그렇지 않다면 혹시 ‘계급’이나 ‘푸로레타리아’와 같은 단어들이 그 불온함을 잃고 잡지 안에서 온건하게 통용되던 방식이 있었을까?

한편 기호의 종류는 달라도 ‘사람’, ‘생각’, ‘문제’ 등의 추상명사들은 반복되고 있는데 이 단어들을 제외하면 좀 더 의미론적으로 변별력 있는 단어들도 찾을 수 있다. 예를 들면 위에 언급한 정치사회 관계어들 (‘주의’, ‘운동’, ‘재벌’, ‘계급’, ‘프롤레타리아’, ‘정치’, 등—주로 여러 복자 중 ‘XX’ 앞뒤로 등장한 단어들)도 있고, ‘사체’, ‘구타’ 등, 사건-사고에 관계된 단어도 있어 호기심을 자극한다. 그러나 이보다는 ‘얼굴’, ‘세월’, ‘결혼’, ‘사랑’, ‘연애’, ‘실수’ ‘마누라’ 등 일상어 근처에 복자가 위치한 것이 궁금증을 자아낸다. 어떠한 이유에서 이와 같은 단어에 복자를 넣어야 했을까? 이외에 ‘신문’, ‘병원’, ‘학교’, ‘은행’ 등 단어 앞에 놓인 기호가 고유명사일 가능성이 큰 단어들도 보인다.

이와 같은 단어들을 대하다 보면 연구의 미진함이 느껴지는데, 앞으로 기호 앞에 놓인 단어의 종류와 뒤에 놓인 단어를 분류함으로써 고유명사로 예상되는 기호와 복자로 예상되는 기호를 구분할 필요가 있다. 또한, 아날로그 텍스트에는 존재하지 않았던 전자텍스트에서만 사용된 기호도 가려낼 필요가 있다. 예를 들면, 이 리스트의 ‘***’ 칼럼에 등장한 ‘鹽原局長’라는 이름과 직함 (중일전쟁 이후 황민화정책 실시한 조선총독부 학무국장 시오바라 도키사부로 鹽原時三郎 국장)은 복자의 의미보다는 원문 판독 불가를 표현하기 위해 전자텍스트에 사용된 기호이다.⁸⁶⁾ 발견의 놀람이 연구의 미진함으로 귀결된 경우로, 앞으로 계속 이러한 예는 리스트에서 지워가야 할 것이다. 이와 더불어 『별건곤』에서 찾은 기호와 공기어와의 관계를 『삼천리』의 것과 비교하면 자주 등장하는 관계의 의미를 좀더 변별적으로 살펴볼 수 있을 것으로 기대한다.

위의 <표 4>에서 본 단어들이 우리가 살펴봐야 할 혹은 연구대상에서 제거해야 할 단어의 폭을 보여 준다면, 그 폭은 아직 넓다. 복자후보 기호를 중심으로 앞뒤 두 단어를 기준으로 모은 것이기 때문이다. 그 폭을 유지하되 일차적으로 어떤 단어에 초점을 맞추어 복자후보가 나온 문맥을 찾아보아야 할까? 이 질문에 대한 대답을 위해 복자후보를 중심으로 앞뒤 한 단어로 제한하여 고빈도어의 표를 만들어보았다. 그 결과는 <표 5>와 같다.

86) “제6문-(****과 東京에서는 内地****과 *** 관계가 있는가요. 서로 有*的****조직체인가요. 鹽原局長(國民總動員朝鮮聯盟)-질문의 **는 六****에 밋쳤으나 ***된 질서있는 질문이기예=이하 4줄 판독 불능)<37> 맨 끝으로부터 順次 답변하리다.” 「징병·의무교육·총동원 문제로 군부와 총독부 당국에 民間有志가 문의하는 會」, 『삼천리』 11.7, 1939, 30~39면.

〈표 5〉 복자후보 앞뒤로 출현한 공기어

별건곤

복자	공기어
XX	사람, 학교, 우리, 재벌, 조선
XXX	사람, 사체, 우리, 생각, 디방
**	사람, 우리, 지불, 고려, 대립
***	면회, 人的祖先, 사람, 원시인, 전설
OO	사람, 학교, 우리, 주인, 선생
OOO	사람, 선생, 처녀, 우리, 광장

삼천리

복자	공기어
XX	우리, 朝鮮, 사람, 운동, 주의
XXX	조선, 문제, 사람, 農民同盟, 우리
**	사람, 관계, 日本, 朝鮮, 까닭
***	관계, 사람, 內地, 東京, 까닭
OO	부대, 은행, 우리, 사람, 운동
OOO	스칸지나비아, 근무, 中佐, 投資, 李應俊

‘사람’, ‘우리’, ‘학교’, ‘조선’과 같은 의미 폭이 상당한 단어들이 먼저 눈에 띄지만, 『별건곤』에서는 ‘재벌’, ‘사체’, ‘디방[지방]’, ‘지불’, ‘원시인’, ‘전설’과 같이 좀더 의미망이 좁은 단어들, 『삼천리』에서는 ‘운동’, ‘주의’, ‘農民同盟’, ‘스칸지나비아’, ‘투자’, ‘李應俊’ 단어들이 보인다. 먼저 이 단어들이 사용된 문맥을 먼저 조사하고 차후에 각 잡지에서 문장 단위로 등장한 복자후보의 공기어들이 실제로 사용된 문맥을 살펴볼 것이다.

공기어 기반의 단어 군집을 통해 맥락을 살펴보는 것은 문헌이나 문장과 같은 분석단위에 함께 나온 단어들을 분석 대상으로 한다. 분석단위 내에서 함께 나온 단어를 손쉽게 파악할 수 있지만, 문장이나 문헌이 길 때 한 분석단위 내에서도 거리가 멀어 의미상으로 연관성이 떨어지는 단어까지 포함할 가능성이 크다. 이러한 단점을 개선하기 위해 특정 단어의 앞뒤 단어들과 그 단어들의 순서를 반영하여 유사단어를 추출하는 word2vec⁸⁷⁾이 개발되었다. word2vec은 문장에 등장한 단어들의 앞뒤

문맥의 단어를 반영하여 단어를 수치화하는 방법으로, 비슷한 의미가 있는 단어는 비슷한 문맥에서 등장한다는 가설에 기반을 두어 단어들을 표현한다. 딥러닝 이전에는 잠재의미분석(latent semantic analysis)이나 토픽모델링으로 잘 알려진 잠재 디리슈레 할당(latent dirichelet allocation)을 통해 문헌을 벡터화하는 시도들이 있어 왔지만, word2vec은 네거티브 샘플링을 통해 학습의 효율을 높이고 수치화된 벡터에 단어의 의미를 내재하도록 했다.

본 연구에서는 복자를 포함하고 있는 텍스트를 대상으로 word2vec을 적용하여 단어들을 임베딩하고 유사한 맥락에서 사용된 단어들을 찾아보았다. 텍스트에서 3번 이상 나온 단어를 대상으로 500차원 벡터로 단어를 표현하고 동시 출현 빈도를 측정할 때와 유사하게 앞뒤 문맥단어로 다섯 단어를 사용하여 학습에 사용했다. 이러한 방법을 통해 만든 2,829,796개의 단어 벡터를 대상으로 코사인 유사도를 사용하여 유사한 용법이나 비슷한 위치에 출현한 단어들을 추출하였다. <표 6>은 코사인 유사도 기준의 유사어 상위 20개 단어이다.

<표 6> Word2vec 기준 복자별 유사어 상위 20단어

순위	XX	XXX	**	***	OO	OOO
1	민족운동자	던도부인	언론기관	恢復	금은상	子正後
2	경제투쟁	侍衛	조사표	半公式	무장단	금은상
3	胎內	피고인	割强	朝鮮人勞働者	春桃	春桃
4	前衛分子	상반부	조사	過去運動	子正後	香港
5	푸로레타리아	금은상	윤길호	先決問題	大同團	O君
6	X略	사체	續報	水田地帶	반양제	韓某
7	해계모니	最後的	自O운동	政略上	韓某	閔某
8	大衆黨	난타	爲替關係	管轄權	산일	반양제
9	지배계급	시약시	사실	戰略的	넘버원	銀紙
10	勞資關係	챙피	수험자	全國經濟委員會	松原上等兵	怒罵
11	도시노동자	民籍	원내외	廣東事變	불죽불죽	무장단
12	最後的	피안	고심초사	生産組織	박형	머누님

87) Mikolov, et al., 앞의 논문.

13	全民族的	배반자	大增收	不利益	金少佐	W町
14	治者	법망	광종	反宗教鬭爭	間島領警	大同團
15	경제덕	발견자	정근	疎忽	개자식	姓字
16	欺瞞的	大衆化	馬券	特殊情勢	단출	자씨
17	隊員	米豆	계산	窮地	O君	속장
18	푸로레타리아트	김목사	통계학자	認識	閔某	성연
19	在中	압살	재소	宋子文氏	괴자	針小棒大
20	인테리켄차	구룡산	발표	施行區域	針小棒大	計算書

위의 결과를 보면, 앞서 공기어 분석에서 보았던, ‘사람’, ‘생각’, ‘문제’ 등의 일반 추상명사보다는 좀더 맥락이 확실한 ‘前衛分子’, ‘大衆黨’, ‘도시노동자’, ‘自O운동’, ‘던도부인’과 회고, 수기, 혹은 소설의 주인공 이름 (‘윤길호’, ‘春桃’ 등)까지 등장한다. 이 주인공은 어떠한 인물이길래 XX 나 OO 등이 반복 사용된 것일까? 검열 항목에는 사회의 안녕질서(치안)를 저해하는 행위와 더불어 풍속괴란이 있는데, 위에 리스트 된 단어 중 치정이나 외설 같은 풍속괴란 항목 때문에 복자로 표현해야만 했던 사례도 있지 않을까?⁸⁸⁾ 앞으로 문장단위, 또한 복자후보 앞뒤로 등장한 단어들과 함께 word2vec에서 찾아낸 어휘들을 참고하여 이 단어들이 등장한 기사를 꼼꼼히 읽고 기호가 사용된 맥락을 파악해 볼 것이다.

VIII. 결론을 대신하여

본 연구는, 국문학 내 문학사회학의 흐름을 거시적으로 조망하는 한편, ‘멀리서 읽기’를 그 지류로 놓고 앞으로 새롭게 개척할 분야와 방법론을 예비해 보고자 하였

88) 이에 관련하여 한 가지 흥미로운 사례는, 평신도가 쓴 “평신도의 수기,” 『별건곤』 6, 1927라는 글이다. 이 기사는 word2vec이 찾아낸 “던도부인”(전도부인)이라는 단어와 관련 깊다. 이 수기는 주인공의 이름을 복자처리하고(“XX” 혹은 “XXX”) 교회 안에서 발생한 치정관계를 소설식으로 기술한다. 외설적인 장면도 삽입되어 있는데, “XX” 나 “OO”같은 복자가 기호 뒤의 텍스트를 감추는 기능보다 독자의 시선을 끄는 데 더 사용된 것이 아닌가 하는 추측도 불리워진다.

다. 신문이나 잡지 매체를 거친 문학 생산방식의 이해는 서구문학과 크게 나뉘는 국문학 생산방식의 한 성격으로 이해될 만하며, 이와 연결되는 식민지 시기 검열연구는 제국과 식민 사이의 비균질적 지식장 형성과 지식장 내 운동을 파악할 수 있는 중요한 영역이다. 본 연구는 기존의 매체연구와 검열연구의 성과를 받아들이고 이를 연결하는 가운데 워드 임베딩이나 워드투벡과 같은 전산분석을 활용하여 복자 후보 기호의 공기어를 찾아보고자 하였다.

후자의 작업에 관한 지금까지의 결과는 초보적인 수준으로, 분석보다는 관찰에 가깝다. 본 논문을 시작하며 상정했던 복자의 앞뒤에 등장하는 어떤 어휘적 패턴이 있는가에 관한 질문에는 대답하지 못했다. 이에 의미 있는 대답을 하기 위해서는 먼저 아날로그 텍스트와 디지털 텍스트의 복자 표기에 차이가 있는지 먼저 확인해야 한다. 『별건곤』과 『삼천리』 텍스트 전체를 비교할 수는 없으므로 여러 전산 분석을 통해 드러난 주요 키워드를 중심으로, 복자기호 표현과 역할을 확인할 필요가 있다. 또한, 단어 추출과 어휘 분석에 있어 더 심도 있는 방법을 찾을 필요가 있다. 예를 들면, 단어의 형태적인 특징을 고려하여 유사한 단어들을 추출⁸⁹⁾하는 FastText⁹⁰⁾이나, 문헌 단위에서의 공기 관계를 고려한 Glove⁹¹⁾ 등 word2vec의 단점을 보완한 다양한 모델을 적용하면—복자의 주변에 나타난 공기어의 패턴과 그 잠재적 의미를 파악할 수 있을 것이다. 이렇게 더 정교한 방식으로 복자 관련 키워드의 범위를 줄이고 이들이 등장한 기사의 문맥을 살펴서 복자의 기능을 두 잡지의 매체 전략의 입장에서 이해해 볼 수 있는지가 다음 논문의 주제가 될 것이다. 본 논문에서는 이를 위한

89) 언어의 형태학적(morphological)인 특성을 반영하지 못하는 word2vec과는 달리 fastText는 단어를 n-gram으로 분할한 문자(character)의 결합으로 취급한다. 따라서 ‘푸로레타리아’와 ‘푸로레타리아트’는 word2vec에서는 전혀 다른 단어로 인식하지만, fastText에서 tri-gram(3글자)으로 처리했을 때, 아래처럼 4개의 공통된 tri-gram이 발견되어 유사한 단어로 인식한다.

- 푸로레타리아 : 푸로레, 로레타, 레타리, 타리아
- 푸로레타리아트 : 푸로레, 로레타, 레타리, 타리아, 리아트

90) P. Bojanowski, E. Grave, A. Joulin, & T. Mikolov, “Enriching word vectors with subword information”, *Transactions of the Association for Computational Linguistics* Vol. 5, 2017, pp. 135~146.

91) J. Pennington, R. Socher, & C. D. Manning, “Glove: Global vectors for word representation”, *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532~1543.

예비단계로, 복자 후보에 몇 가지 전산 분석을 적용한 뒤 얻은 결과를 관찰하고, 어떤 질문이 의미가 있을 것인지 타진해 보았다.

복자의 공기어 분석과 같은 경험적 연구가 아직 미흡한데도 국문학 내 문학사회학적 흐름의 거시적 흐름과 연결한 이유는, 연구방법사의 기술에서 오는 추상적 이해를 보완하고 비슷한 영역에서 겪을 수 있는 연구상의 제반 문제에 대한 해결 과정을 참고로 남기기 위함이다. 이러한 자기 발견적 기록이 앞으로 등장할 새로운 문학사회학 연구의 시행착오를 줄이고 대안적 방법을 찾아가는 데 조금이나마 이바지하기를 바란다.

참고문헌

- 검열연구회, 『식민지 검열-텍스트, 제도, 실천』, 소명출판, 2011
- 골드만 저, 오생근 역, 「소설사회학을 위한 서론」, 유종호 편, 『문학 예술과 사회 상황』, 민음사, 1979
- 권영민, 『한국 현대소설 100년』, 전 2권, 동아출판사, 1995
- 권영민 편, 『한국 근대 문인 대사전』, 아세아문화사, 1990
- _____, 『한국 현대 문인 대사전』, 아세아문화사, 1991
- _____, 『한국 현대문학 작품 연표』, 전 2권, 서울대학교 출판부, 1998
- 김병철, 『한국근대번역문학사 연구』, 을유문화사, 1975
- 김영민, 『한국근대소설사』, 솔, 1997
- 김현, 『한국문학의 위상(1983)/문학사회학』, 문학과지성사, 1991
- 로베르 에스카르피 저, 민희식, 민병식 공역, 『문학의 사회학』, 을유문화사, 1983
- _____, 임문영 역, 『책의 혁명』, 보성사, 1985
- 백철, 『조선신문학사조사』, 전 2권, 수선사, 1947; 백양당, 1949
- 송인재, 「신기한 인문학」, 『세계 디지털 인문학의 현황과 전망』, 커뮤니케이션북스, 2019
- 스티븐 제이 굴드 저, 이명희 역, 『폴하우스』, 사이언스북스, 2002
- 이선영, 『한국문학논저 유형별 총목록』, 총 7권, 한국문화사, 1990~2001
- _____, 『한국문학의 사회학』, 태학사, 1993
- 이재옥, 『조선시대 과거 합격자의 디지털 아카이브와 인적 관계망』, 보고서, 2018
- 정근식, 한기형, 이해령, 고노 겐스케, 고영란, 『검열의 제국』, 푸른역사, 2016
- 정선태, 『개화기 신문논설의 서사수용양상』, 소명출판, 1999
- 정진석, 『극비 조선총독부의 언론검열과 탄압』, 커뮤니케이션북스, 2007
- 천정환, 『근대의 책읽기-독자의 탄생과 한국 근대문학』, 푸른역사, 2003
- 최동호, 최유찬 외 편저, 『한국 근대잡지 소재 문학텍스트 연구』, 총 4권, 2012~2013
- 프랑코 모레티 저, 이재연 역, 『그래프, 지도, 나무』, 문학동네, 2020
- 한기형, 『식민지 문역-검열, 이중출판 시장, 피식민자의 문장』, 성균관대학교 출판부, 2019
- _____, 『한국근대소설사의 시각』, 소명출판, 1999
- 한만수, 『허용된 불온-식민지 시기 검열과 한국문학』, 소명출판, 2015
- Franco Moretti, *Distant Reading*, London: Verso, 2013
- Pascale Casanova, *The World Republic of Letters*, 1999, translated by Malcolm DeBevoise, Cambridge: Harvard University Press, 2004
- Priya Joshi, *In Another Country: Colonialism, Culture, and the English Novel in India*, New York: Columbia University Press, 2002

- 강용훈, 「월평의 형성과정과 월평방식의 변화 양상」, 『한국문예비평연구』 34, 2011
- 김바로, 「역사기록의 전자문서 편찬방법 탐구-역사요소를 중심으로」, 『열상고전연구』 51, 2016
- 김용수, 「세계문화와 디지털 인문학 방법론-한국 학계의 모레티 연구」, 『비평과 이론』 24.3, 2019
- 박현호, 손성준, 「한국 근대문학 검열연구의 통계적 접근을 위한 시론-『조선출판경찰월보』와 식민지 조선의 구텐베르크 은하계」, 『외국문학연구』 38, 2010
- 이상경, 「『조선출판경찰월보』에 나타난 문학작품 검열 양상 연구」, 『한국근대문학연구』 17, 2008
- 이재연, 「작가, 매체, 네트워크-1920년대 소설계의 거시적 조망」, 『사이』, 2014
- _____, 「키워드와 네트워크-토픽 모델링으로 본 『개벽』의 주제지도 분석」, 『상허학보』 46, 2016
- 이재옥, 「조선시대 문무과 재급제 현황과 재급제자 조사」, 『장서각』 32, 2014
- 이혜령, 「식민지 검열과 ‘식민지-제국’ 표상-『조선출판경찰월보』의 다섯 가지 통계표가 말해주는 것」, 『대동문화연구』 72, 2010
- 전도현, 「한국 근대잡지 소재 문학텍스트에 대한 통계적 분석」, 『한국학연구』 47, 2013
- 전봉관, 김병준, 이원재, 「문예지를 매개로 한 한국 소설가들의 사회적 지형」, 『현대소설연구』 61, 2016
- 정근식, 「식민지적 검열의 역사적 기원-1904~1910」, 『사회와 역사』 64, 2003
- 정근식, 최경희, 「도서관의 설치와 일제 식민지출판경찰의 체계화, 1926~1929」, 『한국문학연구』 30, 2007
- 정재윤, 모경현, 서승완, 김창엽, 김해동, 강필성, 「워드 임베딩과 단어 네트워크 분석을 활용한 비지도학습 기반의 문서 다중 범주 가중치 산출-휴대폰 리뷰 사례를 중심으로」, 『대한산업공학회지』 44.6, 2018
- 최지명, 「기계학습을 이용한 역사 텍스트의 저자판별-1920년대 『개벽』 잡지의 논설 텍스트」, 『언어와 정보』 22.1, 2018
- 한기형, 「‘법역’과 ‘문역’-제국 내부의 표현력 차이와 출판시장」, 『민족문화사연구』 44, 2010
- _____, 「‘이중출판시장’과 식민지 검열-‘토착성’이라는 문제의식의 제기」, 『민족문화사연구』 57, 2015
- 한만수, 「강경에 ‘소금’의 복자 복원과 복한 ‘복원’본의 비교」, 김인환 외 편, 『강경에, 시대와 문학-강경에 탄생 100주년 기념 남북한 공동논문집』, 랜덤하우스코리아, 2006
- 허 수, 「‘네트워크 분석을 통해 본 1980년대 ‘민중’-『동아일보』의 용례를 중심으로」, 『개념과 소통』 18, 2016

- _____, 「어휘연결망을 통해 본 ‘제국’의 의미-‘제국주의’와 ‘제국’을 중심으로」, 『대동문화연구』 87, 2014
- D. F. McKenzie, “The Book as an Expressive Form”, *Bibliography and the Sociology of Texts*, London: Cambridge, 1999
- Dora Zhang, “Literature and Money: Studying What You Love Without Being Exploited”, *Chronicles of Higher Education*, 2015
- J. Berenike Herrmann, “In a Test Bed with Kafka. Introducing a Mixed-Method Approach to Digital Stylistics”, *Digital Humanities Quarterly* Vol. 11. No. 4, 2017
- Jae-Yon Lee, “Before and after the “Age of Literary Coteries: A Diachronic Analysis of Writers’ Networks in Korea, 1917~1927”, *Korea Journal* Vol. 57. No. 2, 2017
- James F. English, “Everywhere and Nowhere: Sociology of Literature after the ‘Sociology of Literature’”, *New Literary History* 41, 2010
- Jeffrey. Pennington, R. Socher, & C. D. Manning, “Glove: Global vectors for word representation”, *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014
- Katherine Bode, “The Equivalence of ‘Close’ and ‘Distant’ Reading; or, Toward a New Object for Data-Rich Literary History”, *MLQ*, 2017
- Kyeong-Hee Choi, “Impaired Body as Colonial Trope: Kang Kyŏng’ae’s ‘Underground Village’”, *Pubic Culture* Vol. 13. No. 3, 2001
- Mike Frangos, “The End of Literature: Machine Reading and Amitav Ghosh’s The Calcutta Chromosome”, *Digital Humanities Quarterly*, Vol. 7. No. 1, 2013
- Piotr Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information”, *Transactions of the Association for Computational Linguistics* 5, 2017, pp. 135~146.
- Rachel Sagner Burma and Laura Heffernan, “Search and Replace: Josephine Miles and the Origins of Distant Reading”, *Modernism/modernity Print Plus* Vol. 3. Cycle 1, 2018
- Sean Latham and Robert Scholes, “The Rise of Periodical Studies,” *PMLA* Vol. 121. No. 2, 2006
- S. Jänicke, G. Franzini, M. F. Cheema1 and G. Scheuermann, “On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges,” *Eurographics Conference on Visualization (EuroVis)*, 2015
- Stephen Marche, “Literature Is not Data: Against Digital Humanities”, *Los Angeles Review of Books*, 2012

Ted Underwood, “A Geneology of Distant Reading”, *Digital Humanities Quarterly* Vol. 11. No. 2, 2017

Timothy Brennan, “The Digital Humanities Bust,” *The Chronicle Review*, 2017

Tomas Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality”, *Advances in neural information processing systems*, 2013

류인태, 「지암일기-데이터로 다시 읽는 조선시대 양반의 생활」, <http://jiamdiary.info>

Placing Distant Reading in Korea's Literary Sociology — A Way to Initiate a New Method in Censorship Studies

Lee, Jae-yon · Jeong, Yoo-kyung

Scholars of English literature in the U.S. have explored a new quantitative analysis called “distant reading” as a way to link the tradition of literary sociology and the emergent field of digital literary studies. Intrigued by such a methodological trend overseas, this paper aims to newly locate the status of digital literary studies in the genealogy of literary sociology in Korea and revisit the old questions in censorship studies with new approaches. This objective is structured in two parts. First, we trace the shifting focus of literary sociology since the 1980s in Korea, from literary productions to periodical studies, censorship studies, and digital literary studies. After tracing this unfolding, we examine the *fusejis* (i.e. symbols used in the censored words, such as OO, XX) from 1930s Korean magazines through text-mining (word frequency, collocations, and word embedding) and tests whether we can find linguistic patterns before and after the introduction of those symbols. The emphasis is on the first part as the second is still in an experimental stage. Through this attempt we look forward the coming of new approaches in literary sociology in Korea.

Key Words : distant reading, literary sociology, periodical studies, censorship studies, text-mining, modern Korean literature, digital literary studies

